



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# Assessing EEG neuroimaging with machine learning

Andrew X Stewart  
v1.4



Doctor of Philosophy candidate  
Neuroinformatics  
Institute for Adaptive and Neural Computation  
School of Informatics  
University of Edinburgh

# Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(Andrew Stewart)

# Brief Summary

- We examine EEG (scalp electrode) data recorded from human participants while they view simple visual objects.
- As well as conventional EEG analysis, we also explore 'machine learning' methods, where classifiers can be trained on known data, and then label additional new data.
- Using Support Vector Machine (SVM) machine learning, we find that the identity of previously-unseen trials can be correctly determined at suprisingly high accuracy.
- We use Independant Component Analysis (ICA) to tranform the recorded data and reveal possible source generators, and find higher trial classification accuracy using this ICA data.
- We explore the use of this technique in revealing possible information content in visual perception, and suggest how specific use of machine learning can quantify information flow in the brain from EEG neuroimaging data.



# Abstract

Neuroimaging techniques can give novel insights into the nature of human cognition. We do not wish only to label patterns of activity as potentially associated with a cognitive process, but also to probe this in detail, so as to better examine how it may inform mechanistic theories of cognition. A possible approach towards this goal is to extend EEG ‘brain-computer interface’ (BCI) tools – where motor movement intent is classified from brain activity – to also investigate visual cognition experiments.

We hypothesised that, building on BCI techniques, information from visual object tasks could be classified from EEG data. This could allow novel experimental designs to probe visual information processing in the brain. This can be tested and falsified by application of machine learning algorithms to EEG data from a visual experiment, and quantified by scoring the accuracy at which trials can be correctly classified.

Further, we hypothesise that ICA can be used for source-separation of EEG data to produce putative activity patterns associated with visual process mechanisms. Detailed profiling of these ICA sources could be informative to the nature of visual cognition in a way that is not accessible through other means. While ICA has been used previously in removing ‘noise’ from EEG data, profiling the relation of common ICA sources to cognitive processing appears less well explored. This can be tested and falsified by using ICA sources as training data for the machine learning, and quantified by scoring the accuracy at which trials can be correctly classified using this data, while also comparing this with the equivalent EEG data.

We find that machine learning techniques can classify the presence or absence of visual stimuli at 85% accuracy (0.65 AUC) using a single optimised channel of EEG data, and this improves to 87% (0.7 AUC) using data from an equivalent single ICA source. We identify data from this ICA source at time period around 75–125 ms post-stimuli presentation as greatly more informative in decoding the trial label. The most informative ICA source is located in the central occipital region and typically has prominent 10-12Hz synchrony and a  $-5 \mu\text{V}$  ERP dip at around 100ms. This appears to be the best predictor of trial identity in our experiment.

With these findings, we then explore further experimental designs to investigate ongoing visual attention and perception, attempting online classification of vision using these techniques and IC sources. We discuss how these relate to standard EEG landmarks such as the N170 and P300, and compare their use. With this thesis, we explore this methodology of quantifying EEG neuroimaging data with machine learning separation and classification and discuss how this can be used to investigate visual cognition. We hope the greater information from EEG analyses with predictive power of each ICA source quantified by machine learning might give insight and constraints for macro level models of visual cognition.

# Acknowledgements

There are many I owe thanks to in the Neuroinformatics DTC programme at Edinburgh, not least for fostering such an engaging research environment.

I would like to give thanks to my supervisors, Guido Sanguinetti and Antje Nuthmann, for their help. While this work is my own, they provided crucial support. Throughout this thesis, when I speak of the lab ‘we’, I mean to recognise their support, and when I use a first-person ‘I’, I mean to make a claim myself.

I want to give particular thanks to Baris Demiral (currently at NIH) for helping me set up the EEG hardware and the practicalities of running experiments.

Many thanks are also due to Samantha Rogerson, Claire Kaplan and Linda Stewart for help, support and proof-reading throughout the writing process. Any mistakes are my own.

# Contents

<b>Abstract</b>	<b>3</b>
<b>1 Introduction and Core Hypotheses</b>	<b>8</b>
1.1 Theoretical motivation . . . . .	8
1.2 Key Literature . . . . .	9
1.2.1 EEG and visual cognition . . . . .	9
1.2.2 EEG and BCI . . . . .	11
1.2.3 ICA and EEG data mining . . . . .	12
1.2.4 Comparison with the use of neuroimaging in language classification	12
1.2.5 Learning from EEG . . . . .	13
1.2.6 EEG background . . . . .	14
1.3 Neuroimaging and vision . . . . .	17
1.3.1 Relating neuroimaging to the computational neuroscience of vision	17
1.3.2 EEG features and ICA . . . . .	18
1.4 Support vector machines classifiers . . . . .	20
1.5 Online analysis investigation . . . . .	21
1.5.1 Visual reconstruction from fMRI . . . . .	21
1.5.2 Why examine attention? . . . . .	22
1.6 Outline of Core Hypotheses . . . . .	22
1.7 Thesis outline . . . . .	23
<b>2 Methodology</b>	<b>25</b>
2.0.1 Experimental approach . . . . .	25
2.0.2 Extracting relevant features from EEG data . . . . .	25
2.0.3 Use of ICA . . . . .	26
2.0.4 Use of Support Vector Machines (SVM) . . . . .	28
2.0.5 A note on chance and baselines . . . . .	32
2.1 Experimental Methods 1 - Visual Classification I . . . . .	34
2.1.1 Participants . . . . .	34
2.1.2 Materials and design . . . . .	34
2.1.3 Apparatus . . . . .	34
2.1.4 Experimental procedure . . . . .	34
2.1.5 EEG recording . . . . .	35
2.2 Experimental Methods 2 - Online Analysis . . . . .	38
2.2.1 Use of the online attention interface . . . . .	38
2.2.2 Motivation . . . . .	39
2.2.3 Designing the online experiment . . . . .	40
2.2.4 EEG recording environment . . . . .	41
2.2.5 Materials and design . . . . .	42

2.2.6	Experimental procedure . . . . .	44
2.2.7	Online analysis, Stage I differences from Visual Classification I .	47
2.2.8	Neurofeedback setup for stage III . . . . .	48
2.2.9	Online system implementation . . . . .	48
2.2.10	Advantages of online . . . . .	51
<b>3</b>	<b>Results</b>	<b>53</b>
3.1	Results of Visual Classification I . . . . .	53
3.1.1	Hypotheses tested here . . . . .	53
3.1.2	Classification of visual stimuli . . . . .	54
3.1.3	Classification using ICA data . . . . .	57
3.1.4	Percentage correctly classified . . . . .	58
3.1.5	EEG ERP and IC details . . . . .	58
3.1.6	Scalp EEG data with IC artefacts removed . . . . .	58
3.1.7	More precise temporal profiling with a sliding analysis window .	60
3.1.8	Properties of the ICs . . . . .	61
3.1.9	Classification performance on each image . . . . .	63
3.1.10	Single Classification of 'Any Object Present' vs 'All Objects Absent'	63
3.1.11	Object category classification . . . . .	66
3.1.12	Classification across subjects . . . . .	69
3.1.13	Cluster properties . . . . .	69
3.2	Visual classification discussion . . . . .	72
3.2.1	Improving accuracy with selected input . . . . .	72
3.2.2	The use of ICA . . . . .	73
3.2.3	Visual object processing . . . . .	74
3.2.4	Cognitive implications and future work . . . . .	75
3.2.5	Limitations . . . . .	76
3.2.6	Visual Classification summary . . . . .	77
3.3	Results of Online Analysis experiments . . . . .	78
3.3.1	Stage I – training 'one-vs-many' object models . . . . .	78
3.3.2	Stage II – classifying visual with and without attention . . . . .	79
3.3.3	Classification accuracy on ambiguous images, with and without attention . . . . .	80
3.3.4	Attention-specific classifier properties . . . . .	82
3.3.5	Stage III – conscious modulation of attention-dependant compo- nents with feedback . . . . .	83
3.4	Examining Online Analysis . . . . .	84
3.4.1	Online optimisation . . . . .	84
3.4.2	Attention differences . . . . .	84
<b>4</b>	<b>Discussion</b>	<b>86</b>
4.1	Hypothesis 1 . . . . .	86
4.1.1	Theoretical implications and relation to literature . . . . .	87
4.2	Hypothesis 2 . . . . .	88
4.2.1	Theoretical implications and relation to literature . . . . .	88
4.3	Hypothesis 3 . . . . .	88
4.4	Hypothesis 4 . . . . .	90
4.5	Hypothesis 5 . . . . .	91
4.6	Visual Cognition . . . . .	92
4.6.1	Perception is serial in focus . . . . .	92

4.6.2	A single IC source located occipital-parietal region better predicts object identity . . . . .	92
4.7	EEG processing findings . . . . .	93
4.8	Machine learning classification findings . . . . .	95
4.9	Potential future work . . . . .	98
4.10	Conclusion . . . . .	99
<b>Appendices</b>		<b>100</b>
<b>A</b>	<b>Appendix A - display testing</b>	<b>101</b>
A.1	Monitor testing . . . . .	101
A.1.1	Monitor results . . . . .	104
A.1.2	Monitor discussion . . . . .	109
<b>B</b>	<b>Appendix B - Paper published on offline single-trials classification of EEG and ICA</b>	<b>110</b>

# Chapter 1

## Introduction and Core Hypotheses

### 1.1 Theoretical motivation

Neuroimaging gives a new window into the workings of an awake human brain. It can allow an objective metric of neural activity, and so theories of cognition can be probed in more detail, tested and informed. Electroencephalography (EEG) measures  $\mu V$ -scale changes in the electric field recorded the scalp, which is sensitive to coalitions of many neurons changing their oscillatory activity (Niedermeyer and Da Silva, 2005). This is very fast and so gives EEG millisecond temporal resolution – much greater than that of fMRI neuroimaging.

The most prevalent research use of EEG has been in tracking event-related potentials (ERPs), typically taken to be the change in the voltage reading from a set of electrodes with the onset of some event, averaged over all occasions of that event. This reliably shows effects like the P300 ERP in response to the oddball paradigm (Polich and Kok, 1995), where the ERP response to a novel presented item is notably distinct from that of more routine visual objects presented to the participant. The use of ERP alone has been criticised (Rousselle and Pernet, 2011; Gaspar et al., 2011) for hiding effects in grand averaging and not using all available information.

Recent advances in computer science, particularly machine learning, allow alternate analysis methodologies that might allow more information to be extracted (Makeig et al., 2004). Independent component analysis (ICA) has been used as an artefact rejection technique for EEG data (Delorme et al., 2007). In essence, it separates the main independent generators of variance within a signal. With three microphones separated spatially in a room, ICA can solve the cocktail-party problem of identifying and isolating three voices speaking together. For EEG then, the detectors of EEG electrodes act as the microphones and the varying electrical pattern as the voices. The ICA components give a metric of source space of generated activity. This is particularly advantageous when trying to classify neural activity, as ICA reduces the dimensionality of signal, making it more amenable to prediction.

An initial motivation for this project was to investigate if – and how well – tools from machine learning could be applied to EEG neuroimaging in the context of visual cognition experiments.

This was partly inspired by the impressive ‘decoding’ of motor intent from EEG data in brain-computer interfaces (BCIs), where classification algorithms can restore some ability to the paralysed, allowing them to move a computer cursor. Can such tools be applied to the domain of visual cognition, to inform us of the sub-steps of visual processing in humans, rather than of motor intent? And if so, what new information can be gleaned? Can this present novel information toward theories of visual cognition?

In working towards a goal of high performance, real time decoding of perceptual processes from EEG neuroimaging, we must establish the nature of the signal and how we can best learn from it. We begin with this in mind and review surrounding key literature, and use this to build well-informed hypotheses. We provide a justification of these and explore how these are to be tested.

## 1.2 Key Literature

### 1.2.1 EEG and visual cognition

As EEG offers neuroimaging with high temporal resolution (recording at millisecond resolution, rather than 1 s resolution of fMRI (Niedermeyer and Da Silva, 2005)), it allows experimental designs that would not be feasible using other tools.

A pertinent example of this is found in the work of Thorpe et al. (1996), where reaction time and EEG response to a visual choice task puts hard limits on the timing of visual processing processes. This was achieved by setting a simple ‘go/no-go’ task in response to the appearance of an animal within a scene flashed before them for 20ms. Mean reaction time was around 420 ms, but EEG signal strongly associated with the participant’s choice was apparent at the 125 ms mark. The authors suggest that this signal is perhaps more likely to represent neural activity associated with the choice, and so the visual processing steps to inform that choice necessarily must precede this.

This point is expanded upon by Johnson and Olshausen (2005), who attempt to separate the decision of the participant to respond from the visual object task that is performed. This is done by using both visual object stimuli and text stimuli, and by varying task difficulty. They conclude that their findings support the conclusion that the early large EEG signature is likely ‘postsensory’ – more related to the decision making cognitive mechanisms than object recognition in and of itself.

Kirchner and Thorpe (2006) adapt this protocol by using a ‘forced choice’ design, where the participant is instructed to move their gaze to the image containing an animal – either to the left or right (see Fig. 1.2.1). They find this saccade is made at very high accuracy and often within 120 ms of stimuli appearing, offering further evidence that the visual processing and decision making must be do-able within this period. They

## Timecourse of information flow in visual areas

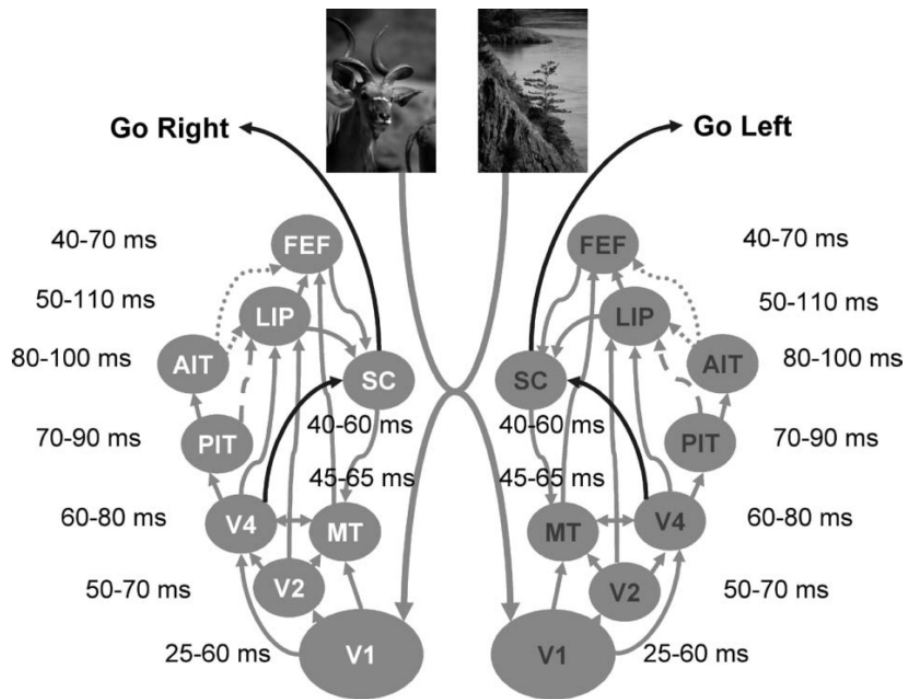


Figure 1.1: Using EEG data a 'forced choice' experimental design, Kirchner and Thorpe (2006) investigate timings and possible localisations in visual processing. High task performance in terms of saccades to the correct direction, given animal image prompt, places limits on the latency of information flow. Times shown are estimates of lower bounds at each anatomical area (represented in shaded bubbles).



suggest a putative information flow within the human neuroanatomy that fits these constraints and discuss the implications for computational models of cognition.

These studies highlight the ability of EEG data to inform investigation of mechanisms of visual cognition. They show that well-designed tasks and EEG recording can provide novel information in visual perception performance and timing. Of course, these studies cannot speak to specific brain source localisation, causality, attention or conscious intent of the participant. These were not within the scope of these studies, but still very much of interest to those wishing to investigate visual cognition.

An experimental design that does address specific localisation, attention and conscious intent is found in Cerf et al. (2010). While working with epilepsy patients with implanted cortical electrodes, a task was designed where multi-unit electrode recording data could be analysed online and was used to classify the participants' response to famous faces. Additionally, in the presence of two overlapping semi-opaque faces, the participant was asked to focus their attention on one target. They found that the participants could modify the activity of temporal lobe neurons to fit the target visual stimuli. They conclude that the participants had voluntary control over the activity of these neural patterns. This design is novel as the ongoing state of the brain changes how a trial proceeds – and so they are able to investigate the degree to which a pattern of activity might be under conscious control in a new way. Unfortunately, there were some limitations with this study, as there were only a small number of patients with suitable electrode implants and the location of these were of course determined by medical need rather than scientific inquiry. The danger and expense of cortical electrode implants limit both the number of patients and the time spent with each one.

From these papers, we can take the lesson that EEG neuroimaging may be a fruitful tool in investigating visual cognition, and high quality processing, source separation and novel online tasks might significantly aid this.

### 1.2.2 EEG and BCI

Research into using EEG recordings and processing online has created the area of EEG brain-computer interfaces (BCIs). These were first investigated both to develop novel hands-free computer input devices and to help those with bodily paralysis (Donoghue, 2008). EEG is the least invasive of these, with no neurosurgery required, yet yields surprisingly high comparative accuracy (Lotte et al., 2007).

There has been considerable interest in developing increasingly successful BCIs, to give better control and improved functionality. This is shown in the multiple iterations of the 'BCI Competition' (Blankertz et al., 2006), with improving performance. These groups typically take EEG data with some timepoints labelled as motor intent (to move a mouse one direction or to pull a lever) and train machine learning models (using some data processing, feature extraction, and classification) to apply a label to new data. These groups are judged by their ability to correctly label previously unseen data, whose labels are withheld by the organisers of the competition. This allows a better

test of the entrants methodology.

One approach is to examine the relative strengths of dominant features observed in EEG data and those known to be functionally associated with motor intent, such as the Mu rhythm, as in Li et al. (2010). This has some advantage of simple, quick training and more 'analogue' cursor movement. A more typical approach is to design artefact-rejection, feature extraction and then train SVM (support vector machines) to classify using this data. This does require more training time, but is used successfully (Blankertz et al., 2006) and more recently in Wang et al. (2012) where the authors suggest greater robustness in that approach.

The great advance in EEG BCI performance indicates that there is a great amount of relevant neural data to be extracted from EEG in this manner, and this suggests that this could also be applied to non-motor investigation.

### **1.2.3 ICA and EEG data mining**

The performance of EEG BCIs has been partly due to the ability to extract relevant 'features' from EEG data. For new tasks, these features will be as-yet unprofiled. Independent Component Analysis (ICA) is a machine learning technique that performs 'blind source separation' (Bell and Sejnowski, 1995). A representation of the data is sought where components are maximally independent, and this is arguably a natural description of the data (Barber, 2011). ICA has been used extensively in recent years in processing EEG data (Luck, 2005), allowing researchers to sidestep, to some extent, solving the intractable 'inverse problem' of computing actual generating sources of EEG activity. It is typically used to identify and reject artefacts from muscle movements, which would otherwise contaminate neural data within the EEG signal (Hyvärinen and Oja, 2000), but has also more recently been used to profile the nature of the IC sources themselves (Onton et al., 2006).

This approach, along with others, has prompted researchers towards a more 'data-driven' EEG data mining methodology (Makeig et al., 2004), rather than the traditional ERP peaks or power shifts that were often previously reported (Gaspar et al., 2011).

### **1.2.4 Comparison with the use of neuroimaging in language classification**

The use of machine learning tools on neuroimaging data to investigate a commonly-studied cognitive process has a parallel in the study of language representation. Mitchell et al. (2008) use fMRI data in response to nouns and word categories to predict the fMRI that might be produced in response to new, unseen words. With novel feature extraction, Murphy et al. (2009) were able to apply the same to EEG data, showing that activation patterns can be predicted from semantic representations with significant accuracy. The authors then extend this with what they call a 'collection of advanced data-mining techniques' to greatly improve the classification performance. With this,

they had participants silently name images and the single-trial EEG data could be used to correctly identify the category that this image belonged to at around 60–80% (where  $p = 0.05$  chance would be around 50%). A separate group, using a slightly different methodology, improve on this word and category classification using both MEG and EEG data (Chan et al., 2011) and this gives 83% performance at decoding individual words, although they get lower accuracy for binary categories. Bullinaria and Levy (2013) expand on this area by examining likely limiting factors for investigating semantic representations from brain activity. While they find selection of processing and brain activation vector important, they also point to fMRI vectors as a key improvement. They also suggest the complexity of the linear models may be improved, which seems likely given the accuracy found by the SVMs from Chan et al. (2011).

This surprisingly high accuracy of classification of language from EEG shows how EEG data can reveal new information with adequate data mining and machine learning. This gives a fruitful target for investigation which can be profiled in detail.

### 1.2.5 Learning from EEG

One approach to assess the use of these techniques is to use machine learning on EEG data. Training examples from different experimental states can be used to train machine learning classifiers, and the ability to correctly and robustly identify new examples can be assessed.

This is done in the adjacent field of brain-computer interfacing, where EEG signals are used to control motor prostheses (Wolpaw et al., 2000; Donoghue, 2008). For EEG motor prostheses, the experimental states of interest might be presence or absence of prompted hand movement and success of the classifier would be to robustly identify this motor movement in subsequent EEG data.

We hypothesise a similar approach could be applied to EEG data recorded in visual object presentations. Instead of identifying motor movement, the experimental states could be the presence or absence of a given visual object stimulus (e.g. ‘spoon’).

In EEG motor prostheses, there is a desire for accurate and fast classification. Consequently, many analysis methods have been attempted (Sajda et al., 2003; Lotte et al., 2007). Use of data preprocessing and machine learning tools have proved effective in improving the ability to use EEG data to predict movement intent (Müller et al., 2008).

The use of machine learning on visual EEG experiments is not as well studied as EEG data in motor movement. The EEG response elicited from different object stimuli is likely to be less regular, more subtle and of lower signal-to-noise than EEG from different motor movements, but we suggest recent advances from motor brain-computer interfaces might also apply to improving learning from EEG data in visual experiments.

Therefore, we designed a simple visual object presentation experiment where we labelled EEG data (as ‘object stimuli onset’ or ‘object absent’), trained an SVM classifier



Figure 1.2: A willing experimental participant, wearing a cap with 64 active electrodes, connected to Biosemi ActiveTwo (not shown).

with this data and labels, and assessed that model’s accuracy at labelling subsequent unseen EEG data correctly. We suggest that classifier task-labelling accuracy may be used as a metric of task-relevant information in the data used to train the classifier. This allows comparison of EEG data processing methods.

### 1.2.6 EEG background

In order to sensibly use a tool to investigate the brain, we must have some knowledge of this tool. For us to appropriately use EEG neuroimaging, we must have an awareness of what it is that is being recorded.

#### What is EEG?

Electroencephalography (EEG) relies on electrodes on the surface of the head, recording the voltage changes around the order of  $10\mu V$ .

While the electrical rhythms detected in animal brains had been described before, it was German physiologist Hans Berger who first reported human EEG recordings (Swartz, 1998). This was then confirmed by other clinicians and researchers like Adrian and Matthews in 1934 who used it to study epilepsy.

A plausible explanation for what might be generating these observed rhythms is that with the synchronous firing of many cortical neurons, the flow of ions give an electric field that can be detected on the scalp (Luck, 2005). A simple model of a single voltage generator inside the head being observed and recorded by electrodes at the scalp surface is tempting, but that hides the likely underlying complexity of millions of interacting sources of electric field, summing and cancelling each other, at multiple

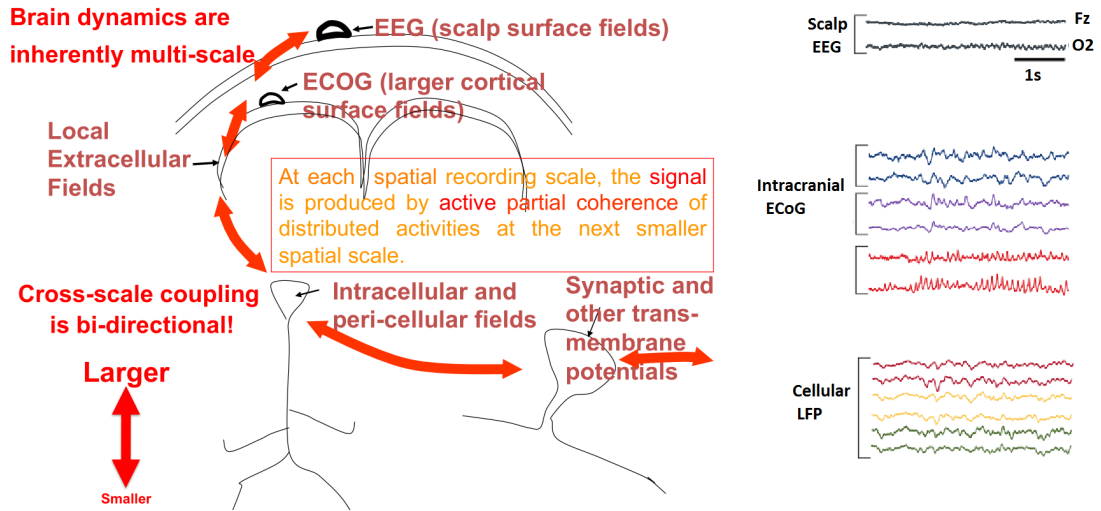


Figure 1.3: A sketch of EEG scalp surface electrodes, placed in comparison with other levels of brain dynamics often investigated. While electrocorticography (ECoG) is recorded within the skull, it does have a high overlap with EEG data. There exist local field potential changes that influence and contribute to firing of neurons, and synaptic events that involve cascades of ions crossing a neurons membrane. (Adapted from (Makeig et al., 2004) and (Buzsáki et al., 2012) - very much not to scale).

scales (see Fig 1.3).

While it is difficult to identify exactly what brain dynamics are represented at these different levels in the brain, some excellent progress has been made, as in (Buzsáki et al., 2012). Even without detailed mechanism of generation, EEG can be empirically observed to reveal information on the brain in states of sleep, epilepsy, surprise (in P300 signal), and other behavioural experiments.

EEG is used extensively in medicine for diagnosing and tracking epilepsy and sleep. EEG has been identified as the best method for profiling epileptic activity in the brain (Noachtar and Rémi, 2009). The extensive use of EEG here, over many years and in the possibility of alternative methods, is evidence for EEG being particularly useful as a method of acquiring information from the brain.

As EEG provides somewhat objective data from the electrical activity at the scalp, it can also be used in psychological and neuroscience experiments in order to probe how this may change in different conditions. An example EEG design is to record the EEG in response to many stimuli visual objects presented to the participants, and then average the voltage across each of the trials. As the electric field is typically somewhat noisy at the level of around  $10\mu V$ , judging unprocessed single trials by eye is not very informative. With even minimal filtering and baseline averaging, several stereotyped features are observed in these kinds of visual prompting experiments (see Fig 1.5).

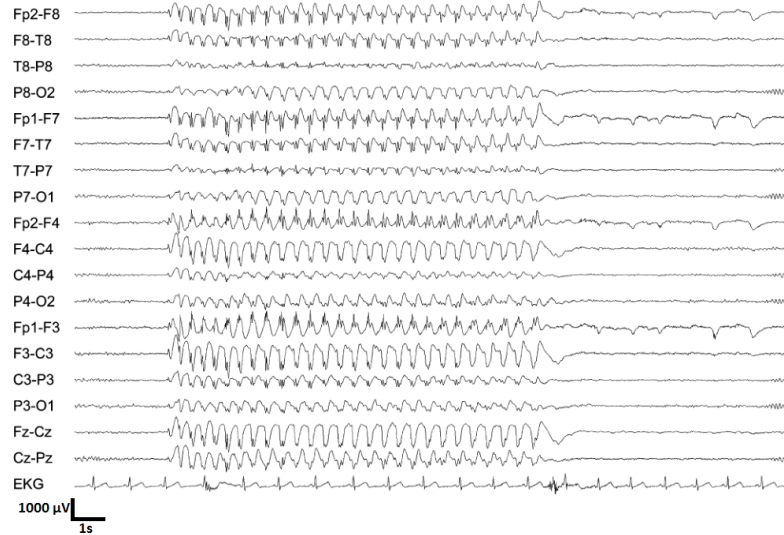


Figure 1.4: Example medical EEG data from 20 continuous seconds from 18 composite channels recording an epileptic patient during a seizure, along with electrocardiograph (ECG/EKG) on lower line. The scale shows around  $1000\mu V$  between line centres. The normal EEG activity is dominated by larger seizure activity after around 3 seconds. Adapted from (Noachtar and Rémi, 2009)

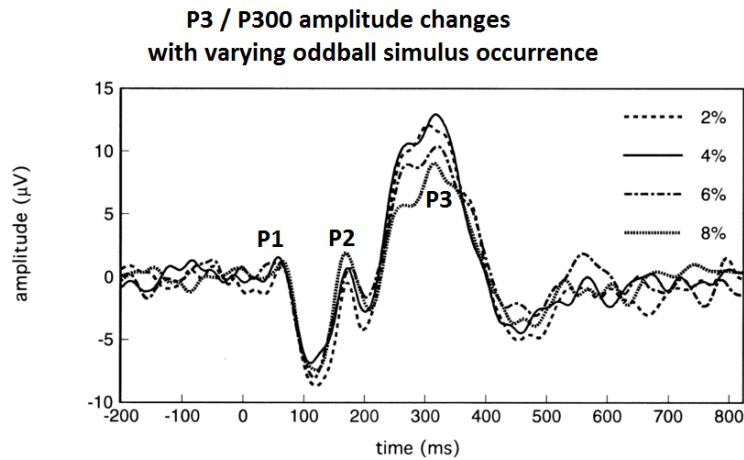


Figure 1.5: Example EEG data from stimulus presentation experiment. Each line in an average activity at electrode Fz across 7 participants with around 500 trials from each. Raw electrode readings were filtered to remove information above 30Hz, and baseline averaged so that prestimulus (as in, 200ms - 0ms) summed to zero. There are three prominent peaks of 'positive deflection', commonly observed at around 100ms, 200ms and 300-450ms, often referred to as P100, P200 and P300, or preferentially P1, P2 and P3 (Luck, 2005). Here, P3 becomes more prominent when the 'oddball' stimulus is more infrequent (Horovitz et al., 2002).

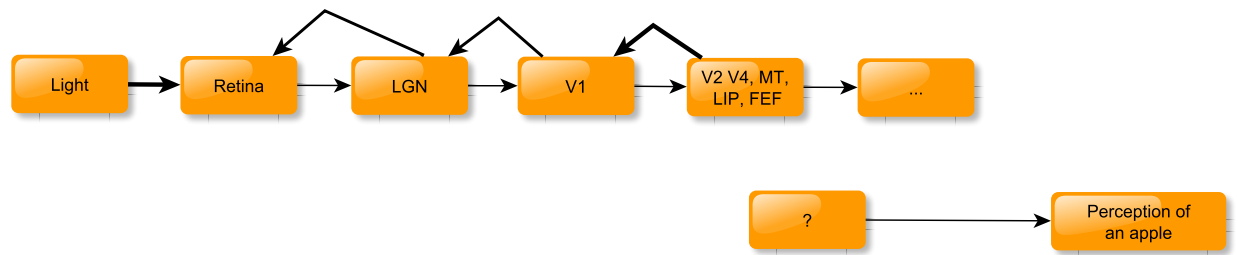


Figure 1.6: A simplistic sketch of information flow in the visual system. The elements listed here are highly likely to be involved in visual perception, but how perception of a visual object arises is much less well understood. Can use of new data mining techniques on neuroimaging data inform any steps in this process?

## 1.3 Neuroimaging and vision

The study of visual perception is a useful and established route to better understanding the conscious brain (Marr, 1982). Reconstruction of visual state has recently been demonstrated (Nishimoto et al., 2011) using machine learning tools on fMRI, another neuroimaging modality, but EEG has high temporal resolution – perhaps greatly more so than is typically used in ERP studies (Makeig et al., 2004; Gaspar et al., 2011). Further studies suggest EEG activity at 150-400ms post visual presentation may contain information of the stimuli (Smith et al., 2012). With ICA isolating EEG sources, a modular detector of distinct visual processes might become more feasible.

### 1.3.1 Relating neuroimaging to the computational neuroscience of vision

The early visual system is comparatively well studied. A simplistic sketch of information flow might look something like Fig 1.6. Work examining the molecular and cellular mechanisms for the excitement of retinal ganglion cells, detection of edges and detection of movement, has revealed an impressive amount about how the early visual system works.

The primary flow of information from photons exciting rhodopsin in retinal cells to primary visual cortex area 'V1' seems clear, while this also heavily modulated by feedback. Fig 1.6 tries to show information flow here in linear fashion, but even in this simplistic sketch, it is important to stress the large effect of feedback from higher to lower areas. We also have some reason to believe that the lower right hand portion (as in Fig 1.6) is correct, as humans frequently report perceiving visual objects. While this is usually dependant on visual stimulation, this is not well defined.

Fig 1.7 points to some likely relations between layers and lobes in vision. How can we examine the activity of these areas while a participant is awake and behaving?

In investigating visual perception, there is an advantage in having participants who

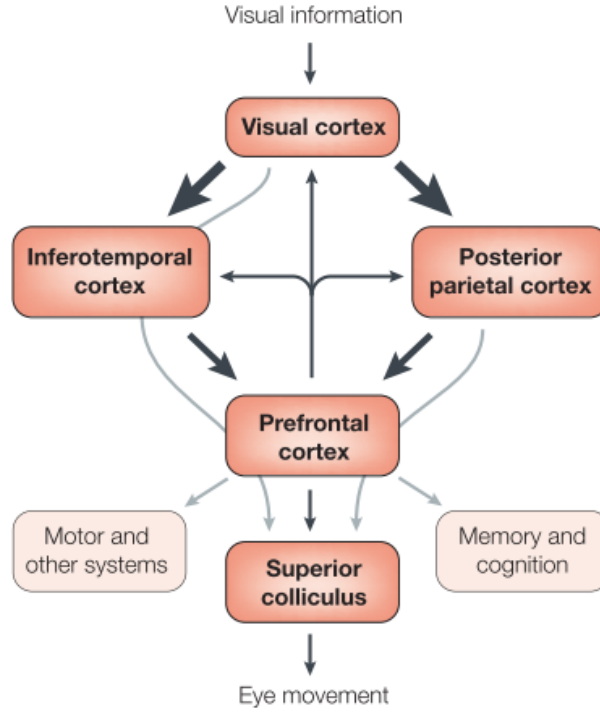


Figure 1.7: Possible information flow in cortical control of visual attention from Itti and Koch (2001). Does the occipital IC we identified track information in V1? Is it susceptible to prefrontal modulation in attention?

can verbally confirm the content of their perception – so humans tend to be better than cats. In addition to the systems handling visual perception, there is also some degree of conscious modulation of visual perception and control of visual attention. Theories of the frontal-cortex driving activity in feedback connections to V1 (Itti and Koch, 2001), and low level saliency biasing cortical activation both explain some findings. We suggest that using new neuroimaging methods, the subprocesses of vision could be better investigated.

### 1.3.2 EEG features and ICA

In the field of machine learning, the input channels of the data that are used to train classifiers are termed 'features'. To improve the accuracy of brain-computer interfaces, a useful technique has been to focus on a small spatial subset of relevant features from EEG. An example of this is identifying 'common spatial features' (Müller et al., 2008; Wang and Jung, 2012). It is well known that these motor signals are highly localised in the motor strip and premotor areas. This made identification of reduced subsets of relevant features tractable (Blankertz et al., 2008). A comparable feature extraction procedure has not yet been accomplished for visual EEG data.

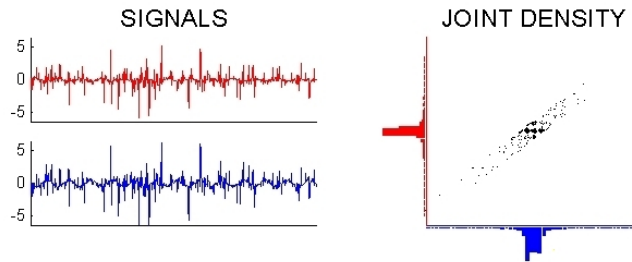
In order to give a comparable decomposition that could be relevant in vision, Inde-



# Independent Component Analysis

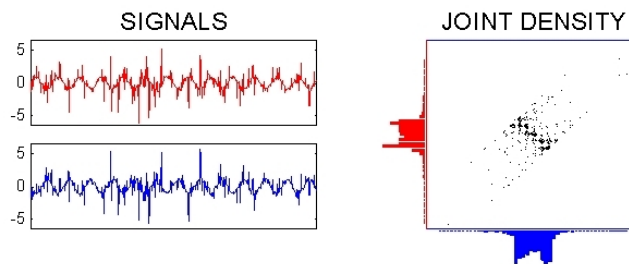
## Steps

0



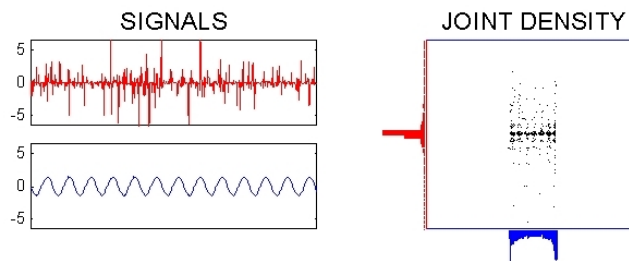
Input signals and density

2



Whitened signals and density

4



Separated signals after 4 steps of FastICA

Credit:  
<http://research.ics.aalto.fi/ica/icademo/>

Figure 1.8: A demonstration of ICA decomposing data into independent sub-parts. Adapted from Hyvärinen (2011) and <http://research.ics.aalto.fi/ica/icademo/>

pendent Component Analysis (ICA) was used on the EEG data in the present study. Each of the IC sources found from the EEG can be considered a reduced subset of the activity in the EEG. In essence, it separates the main independent generators of variance within a signal (Bell and Sejnowski, 1995). An illustrative example of ICA is the ‘cocktail-party problem’ where, given data from three microphones in a room and three overlapping voices, ICA can separate the three individual voices as three distinct sources (Hyvärinen and Oja, 2000). For EEG data then, the detectors of EEG electrodes act as the microphones and the varying electrical patterns as the voices.

ICA is typically used for identifying and removing noisy electrodes, blinks and other artefacts to ‘clean up’ EEG data (Luck, 2005) before proceeding with conventional ERP analysis. Here, we instead use ICA as another way of describing the EEG data, to give subsets of data that may be both more interpretable and give higher classifier performance. The ICA components provide an estimation of possible ‘sources’ of generated activity. This can be particularly advantageous in EEG data analysis, where there is much common signal (and noise) across all channels (Onton et al., 2006). In sum, the identified IC sources may be useful descriptions of subsets of variation within the EEG, as ‘common spatial features’ have been in motor EEG.

## 1.4 Support vector machines classifiers

In experiments here, we sought to determine whether EEG data in our experiment can be automatically classified using machine learning tools. To this end, we used Support Vector Machines (SVMs) – a flexible and powerful statistical learning tool (Burgess, 1998; Cortes and Vapnik, 1995). This technique has given particularly good results in a wide range of domains, including cancer classification (Furey et al., 2000) and face detection (Osuna et al., 1997). In classification of EEG, SVMs have shown good performance in many contexts (Lotte et al., 2007). SVMs were the most commonly used technique for highest accuracy in a EEG classification competition (Blankertz et al., 2006). Thus, we used SVMs as our classifier here.

Specifically, we used SVMs to classify the EEG data into two classes according to the presence or absence of stimulus onset. The underlying principle of SVM classification is to solve a (non-linear) classification problem by transforming it into a linear classification problem in a different, higher dimensional space (or *feature space*).

This is achieved by introducing a non-linear map (*feature map*) into the feature space, which can often be an infinite dimensional space of function. The important aspect of this procedure is that, for many widely used algorithms, one is only interested in the scalar products between pairs of feature vectors; these scalar products can be computed by means of a *kernel function* which depends only on the original (non-transformed) data points. Therefore, the need to work explicitly in the high dimensional space is removed and the feature map is defined implicitly by a choice of kernel function.

A commonly adopted kernel is the so called *Radial Basis Function* (RBF) kernel;

the scalar product of the images of two data points  $x$  and  $y$  under the feature map implied by the RBF kernel is computed as:

$$k(x, y) = \exp[-\gamma \|x - y\|^2] \quad (1.1)$$

where  $\gamma$  is a tunable parameter.

Once a kernel function is selected, the SVM algorithm works by identifying a hyperplane in feature space that separates optimally the two classes in the training data, giving the maximum margin between the images in feature space of the points in the two classes. Often it is desirable to allow a few misclassifications in order to achieve a wider margin of separation; this trade-off is controlled by another parameter, called the training error cost and usually denoted by  $C$ .

## 1.5 Online analysis investigation

If it is possible to predict properties of the visual perceptual state of participants using these machine learning techniques on EEG data, a natural next question is to investigate the limits of this. It appears we have good reason to examine classification of attention in visual cognition.

### 1.5.1 Visual reconstruction from fMRI

Novel neuroimaging work has revealed a surprising depth of information that can be revealed with data processing on visual experiments. Recent work in classification of data from fMRI (see especially Nishimoto et al. (2011) and Miyawaki et al. (2008)) has not only demonstrated impressive results in classifying fMRI data from visual presentation experiments, but also presented insights into cortical representation of visual processes, showing how such information can be inferred from these novel neuroimaging experimental designs.

A point that should be reinforced, and one that Nishimoto et al. (2011) make well, is that "quantitative modelling of human brain activity can provide crucial insights about cortical representations" of visual information. In this way, we build more and more refined models of information processing in the brain by exposing them to empirical results informed by detailed behavioural and neuroimaging experiments.

EEG data – as we are recording here – is quite different from fMRI data processed in those experiments, as EEG deals with electric potential changes recorded at the scalp surface. The ability to resolve spatial information from deeper structures is less in EEG, and so any experimental design using EEG will not be as spatially precise as that in Nishimoto et al. (2011). However, there is also scope for novel information to be garnered from EEG data, from the millisecond timecourse and from the different nature of the emergent electrical potential, and so analysis of EEG data in this context could be quite different and complementary to the existing fMRI work.

Might a similar approach be feasible in EEG, adapting from a machine learning classifier approach here? Could this also give new insights into information representation in the brain? What research areas might this be particularly amenable to investigating?

Directly performing visual reconstruction from EEG data appears quite ambitious. Initially, we suggest an adjacent research target of attempting to use EEG data to track visual perceptive state in terms of object perceived and object attended too.

### 1.5.2 Why examine attention?

Investigating the nature of attentional processes in the brain is an active research area in psychology and neuroscience. Computational models of visual attention consider aspects of the brain and visual stimuli that give rise to some objects being brought into the subjects' attention (Tsotsos and Rothenstein, 2011). Finding evidence towards how eye movements, stimuli, bottom-up and top-down neural processes come together can inform how this works (Itti and Koch, 2001).

A further novel and potentially fruitful approach might be to attempt to track ongoing visual attention in the subject using EEG recording information. Methodologically, I suggest that much of our current lab testing environment, data processing and machine learning setup from the Visual Classification I experiment 2.1 can be adapted to focus on classifying ongoing visual attention.

Towards this, we suggest a final main 'Core Hypothesis' - that high-accuracy classification of EEG data during visual presentation and target prompting can allow testing of how each source may be modulated by visual attention and intent.

## 1.6 Outline of Core Hypotheses

Building from the surrounding literature, we have identified that EEG neuroimaging and new processing methods may be promising in investigating visual cognition. But what core hypotheses would best assess these research avenues? Any core hypotheses should be tractable, demonstrable and quantifiable.

To summarise, our core hypotheses are:

**Core Hypothesis 1** – Machine learning classifiers can be used to predict visual stimuli presence at above-chance accuracy.

**Core Hypothesis 2** – ICA can be used to separate putative EEG sources, and the resultant components can also be used with classifiers to predict visual stimuli presence at above-chance accuracy.

**Core Hypothesis 3** – Additionally, the relative predictive performance of classifiers using the EEG channels and IC source data indicates relative relevant information in each of those sources.

**Core Hypothesis 4** – The properties of these sources, along with their trial-classification performance, can inform and constrain theories of how they may be involved with visual processing and cognition.

**Core Hypothesis 5** – High-accuracy online classification of EEG data during visual presentation and target prompting can allow testing of how each source may be modulated by visual attention and intent.

The first two of these hypotheses should be determined by the classification performance of SVMs trained on the appropriately processed data, judged in area-under-a-ROC-curve, from a simple visual stimuli presentation task. The final point has dependencies of an already developed high-accuracy visual EEG classifier system and more complicated task, as in Cerf et al. (2010).

With these core research hypotheses in mind, we can proceed to specifying methodology to test them.

## 1.7 Thesis outline

### Chapter 1 - Introduction, Background and Core Hypotheses

In this first chapter, an introduction of essential concepts is given, and the core hypotheses examined in this thesis are outlined.

Understanding the mechanisms of vision has been seen as a crucial - and perhaps relatively accessible - window to investigating the mechanisms of cognition (Marr, 1982; Koch, 2004; Zhaoping, 2014). There are competing theories of human visual perception - particularly in attention, and top-down and bottom-up processes.

Separately, there have been advances in ‘data mining’ using computational methods that have shown promise in revealing new insights from EEG data (Makeig et al., 2004). In addition, use of machine learning tools like SVM have shown high performance in labelling EEG data within the area of brain-computer interfaces (Donoghue, 2008).

A systematic review of related work is presented, focussing on data science work on learning from neuroimaging data, particularly in visual processing experiments.

### Chapter 2 - Methodology

In order to assess feasibility of utilising these computational methods and machine learning methods on visual EEG data, a visual object presentation experiment was designed and run. A variety of common objects were shown to participants with EEG being recorded. After processing, SVM models are trained on a subset of training trials of this EEG data. These classifiers were then used to predict the displayed stimuli, given test EEG data. We design, code, and run a new batch of online experiments in which visual attention is first directed to one of two simultaneously shown objects, then models are trained to classify visual attention in real time, with participants given feedback of classifier output.

Here, justification is given as to why these specific methods were chosen and details given on their exact use.

## **Chapter 3 - Results of Classification of Visual Objects from EEG**

I report the experiment results, investigating what properties affect accuracy (such as precise timing and spatial location of training data), and discuss the use of ICA here. This chapter expands on material also published in Journal of Neuroscience Methods (Stewart et al., 2014).

With above-chance accuracy classification of EEG data in visual object trials, this is extended in Chapter 3 to run a new, fully-online experimental design. Object classifiers are trained while the experiment is ongoing. Classification can then be performed on incoming EEG data, and so online feedback can be administered, dependant on properties we detect.

Results are reported from this online classification experiment. We discuss the conscious control of feedback when visual attention is directed, and how this can be indicative of distinct visual processes.

## **Chapter 4 - Discussion**

What has been learned from this project? Given the existing literature and in light of the main results from Chapter 3, what information can we add to help understand visual cognition?

I discuss these questions, use this data to explore the core hypotheses, and point to some additional experiments that appear particularly experimentally amenable to this approach.

## Chapter 2

# Methodology

In this chapter, the selection of specific methods to test the core hypotheses the selection of specific methods, and their use is specified in detail.

### 2.0.1 Experimental approach

In order to test the Core Hypotheses as set out in earlier chapter 1.6, we required specification of simple experiments that could practicably be undertaken to inform them.

An initial requirement was to acquire access to high quality EEG data from experiments with visual stimuli state recorded precisely. This gives training and test data. From existing literature, it appeared that an experimental design close to Thorpe et al. (1996) would allow us to connect to prior work on visual stimuli. This involves presentation of visual stimuli to a human participant while EEG data is being recorded and timing is precisely recorded. We call this stage 'Visual Classification I'.

While recording new data does involve lab time, writing experimental protocols, recording data sessions, and data processing, it does also allow flexibility to record additional data as required. For this reason, we chose to record new data. This was performed in the PPLS EEG recording suite at the University of Edinburgh.

### 2.0.2 Extracting relevant features from EEG data

The nature of EEG recording is noisy, erratic traces at the electrodes, with much shared noise. As seen in the epilepsy recording example (see Fig 1.5), continuous EEG recordings are difficult to read, unless the subject is actively seizing. So how can we make sense of it?

A common analysis of EEG data is to present many experimental trials and average the data for these trials together, so that much random noise is cancelled out, and only the variation that changes in the same way in each trial is conserved. This is then the Event-Related Potential (ERP), showing the average electrical potential around a time-locked event. The P3 is a common example, where the size of the deflection at

around 300ms is increased in trials with less-expected stimuli. Examining ERPs has the advantage of being relatively simple to calculate and visualise, and also being well known among researchers. It also has issues in that much data is thrown away in this averaging, such as any activity not locked to the trial, some of which may well have been informative. Over-reliance on grand-averaged ERPs has been criticised (Rousselle and Pernet, 2011).

An alternative analysis description of EEG data is time-frequency analysis. It can be complementary to ERPs. With this, it is the ongoing strengths of oscillations of different frequencies within the data that is considered. As with Fourier analysis, the signal can be thought of as composed of a sum of oscillations. The prominence of different oscillations can reflect the state of the subject, such as increased  $<4\text{Hz}$  'delta waves' in deep sleep, or increased 10-12Hz 'alpha' when awake and alert. These changes can be monitored over timesteps and displayed as spectro-plots of frequency power across time. While time-frequency analysis can describe spectral patterns over time that would not be seen in the ERP, an ERP-style measure of change in response to a trial can be shown in Event-Related Spectral Perturbation, as shown in top plots of IC data in 3.9.

For both human readability and good machine learning, we want to examine clean, concise data that is not missing important aspects of the data. A cherry-picked data abstraction may unintentionally focus on suboptimal features – if the wrong cherries are picked. An uncommon analysis method is to have ICA separate the potential sources of neural activity and noise, and then use machine learning to classify the trial-predictive information within different data subsets, illustrating what information is where. That is the approach we investigate here.

### 2.0.3 Use of ICA

#### Why investigate ICA?

From existing literature (as discussed in 1.2.3), we found that ICA has been used to good effect for blind source separation for identifying sources of noise within EEG (Hyvärinen and Oja, 2000), allowing this non-relevant data to then be removed. This process of using ICA for EEG data processing has become a well-established technique Makeig et al. (1996), with the developers of the EEGLAB software Delorme and Makeig (2004) being prominent researchers in the area.

ICA takes multi-channel data and returns a transform of this data, where data is rearranged into maximally-independent components. While ICA is a frequently used tool for analysing EEG data, examining the nature of each of these components appears less well studied. There have been some attempts to identify novel targets for study from ICA on EEG data (notably Onton et al. (2006)), but this has not been fully developed.

In the core hypotheses, we wish to examine how use of machine learning tools



for separation and classification might better inform how neuroimaging data relate to visual perception. Core hypotheses 2 considers EEG data 'source' separation via ICA, and assessing the predictive value of information from each ICA component (or 'IC') on a visual task.

As the ICA algorithm attempts to find maximally-independent components to describe the data, the activity of some sources of electrical variation (which may be common across many of the recording electrodes) are arranged into the different ICs. Typically, activity resulting from blinks, eye movements, and muscle movements are captured in prominent ICs (Delorme and Makeig, 2004). These ICs are attributed as blinks, eye movements, and muscle movements partly because expert EEG researchers have observed them and noticed the clear resemblance to known activity of these signals. They have been profiled in time, intensity, phase, location, and spectral power.

Identifying data within ICs as perhaps being a blink works when the generator of that signal is well known. When signal generator activity is not well known, as with many EEG sources within the brain, ICs cannot be tagged so easily. We suggest a tentative hypothesis that there may be ICs that are commonly recorded in EEG and represent unknown, previously poorly-profiled EEG activity that may be less salient but could still be useful in investigating perceptual process in the brain.

Core hypothesis 2 specifies concrete and testable approach towards this. While these 'unknown' origin signal generators can be described in terms of their observed properties (variance over time, intensity, phase, implied location, spectral power, etc.), these alone do not reveal how the activity of these signals may be involved with cognition. In order to begin to answer this, we also quantify these ICs in terms of how informative they are of current visual perceptive state. This is done by training predictive models using data from each IC across different trials, and assessing how well this model predicts the state of new data. This process is described in the SVM methods.

From my reading of the surrounding literature, it is not readily apparent how informative these ICs might be in predicting aspects of visual perception and cognition. Given the large EEG response to visual processing (as in settings in Vanrullen and Thorpe (2001)), I anticipate that there are EEG signatures of visual processing that could identify differential object processing at well above chance. With core hypothesis 2, we investigate which ICs within the EEG data represent this.

With the use of a visual classification task, we have observations of behaviour from a human participant as they observe presented visual stimuli, response data from the participant, simultaneous high-resolution multi-electrode EEG data, and detailed position information of where each recording electrode is in space. This allows acquisition of EEG data (which can be transformed into corresponding IC data) from trials where the participant is shown different stimuli.

## ICA implementation

With the target of ICA source separation from EEG data, we consider possible ways of implementing the ICA algorithm. Thankfully, there are software tools in which the ICA algorithm is already maturely programmed. Using one of these means less development time, increased reliability and repeatability. Two prominent packages are FastICA (Hyvärinen and Oja, 2000) and InfoMax (Langlois et al., 2010).

Calculation of the weights of the sphereing matrix to perform the data  $-i$  ICA transform can be a computationally intensive process, especially for high-resolution, many-electrode EEG data. Tutorial workshop data (from EEGLAB materials) suggest 3 hours on a 2012 3 GHz laptop. FastICA has the advantage of being slightly faster and better established. InfoMax is what is used with the EEGLAB software toolbox ([scn.ucsd.edu/eeglab](http://scn.ucsd.edu/eeglab), from Delorme and Makeig (2004)). With sufficient data, there should not be large differences in approaches Oja and Yuan (2006). As EEGLAB is a mature and well-supported toolbox for an EEG data processing pipeline and ICA, we chose to use EEGLAB InfoMax for generating ICA transform data.

With initial exploratory use, ICA failed to run correctly. This resulted in error messages in Matlab and EEGLAB. This was found to be due excessively large datasets, and was fixed by more stringent data processing. With raw EEG datasets of around 2GB being filtered with a 0.5 – 80 Hz bandpass, downsampled, data partitioned into peri-stimulus trials, and rejections of 'noisy' trials and electrodes, smaller 100MB cleaned datasets were produced. With around 50 recording electrodes, a 256 samples/second sampling rate, and 500 1-second experimental trials, this gives a data matrix of around 50x256x500 datapoints per participant in this partially-cleaned dataset.

Infomax ICA was found to successfully complete on this data, using Matlab R2012a, EEGLAB v13, on a 3 GHz Ubuntu laptop with 16GB RAM within 2hrs. This was subsequently automated in the 'load\_chan\_filter\_script4\_rej.m' Matlab script, which performed this same processing on all participants, running on a University compute server.

### 2.0.4 Use of Support Vector Machines (SVM)

#### Why use SVM?

The area we are investigating is use of analysis methods to learn about the brain from neuroimaging data. While ICA can decompose neuroimaging data into possible sources, it does not identify what each of the sources represent or how they might be involved with cognitive processes. In order to address this, we require a method which takes data from these sources and gives an indication of how much information related to cognitive state is contained within.

Brain-computer interfacing (BCI) is an area where this same process is done (Donoghue, 2008). Many research groups are taking data from EEG and using machine learning models in order to classify movement intent (Blankertz et al., 2006). A main goal of BCIs has been to give those with movement problems another way of interacting with

the world – taking EEG data and producing text output or computer cursor movement. There has been some success in classifying the cognitive states of different intended motor movements using machine learning on EEG data.

Adapting from the success of machine learning in classifying motor intent, we designed experiments to use machine learning to assess visual state.

We chose to use SVM as a machine learning classifier as it had been demonstrated that SVM works as a BCI classifier (Lotte et al., 2007) and does so favourably when compared to other established machine learning techniques (Kronegg J., 2005).

Initial exploration of machine learning method implementation here suggested that deep neural networks or hidden markov network classifiers would also be of interest, but we chose to focus research effort on one machine learning method for the sake of simplicity whilst still investigating these core hypotheses. Given the prior evidence for SVMs working well as a EEG classifier, we chose to stick with SVM as the main machine learning classifier method.

### **SVM implementation**

As we were working in Matlab for handling and processing the EEG data, we sought mature SVM software that could easily work in Matlab. Using the inbuilt 'svmclassify' function in Matlab was not convenient in our setup as it required additional paid toolboxes. A well established and mature SVM interface that is available in Matlab was found in LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, Chang and Lin (2011)).

The precise use of SVM was informed by many sources, particularly a tutorial review from Burges (1998), a guide from the authors of LIBSVM Chih-Wei Hsu, Chih-Chung Chang and Lin (2008), and lectures from Andrew Ng (Ng, 2009). In the context of computational biology, Noble (2006) provides a good primer of SVM use. Correctly partitioning, pre-processing, scaling and scoring are all important in efficient use of machine learning.

In the case of our EEG data, we needed to consider how best to handle the data in order to best classify visual state. In a series of Matlab scripts (see `eegclass.m`) we load partially processed EEG data and write data electrode data to *Xall*. In the case of currently examining ICA data,  $Xall = Xall * icaweights * icasphere$ , as we don not want to keep both large data matrices in memory, nor repeatedly convert them.

Data in *Xall* is a large matrix of perhaps 60 channels \* 300,000 timepoints (at a 256 Hz sample rate and 20 minute recording time,  $\approx 256 * 60 * 20$ ), that is an entire, pre-processed experimental session.

It is necessary to first scale the data within this matrix (Chih-Wei Hsu, Chih-Chung Chang and Lin, 2008). Scaling was applied such that the data was 'normalised' across channels, with each data channel having 0 mean and standard deviation of 1, so the majority of the datapoints now lie within [-1] and [1].

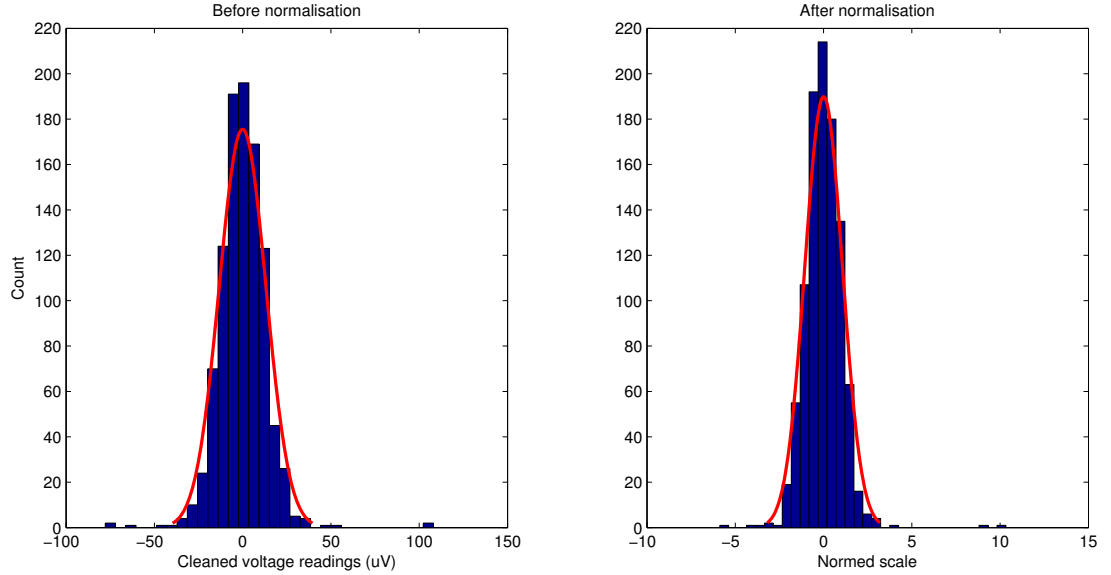


Figure 2.1: Histogram of EEG datapoints before and after scaling each channel for SVM. This illustrates a random subset. Fitted normal density is shown superimposed (red line). Data is 'normalised' with 0 mean and standard deviation of 1.

### Classifier training

Separate classifiers were used for each participant. In the training phase, a 'one-versus-one' SVM model was trained for each of the 50 objects presented, where the labels were 'object stimuli onset' or 'object absent'. The classification task was to best apply this label appropriately to subsequent data, given this training data. That is, rather than using a single multi-class classifier that would give output of 1 of 50 labels, we used 50 binary classifiers each labelling 'object stimuli onset' or 'object absent' for their respective object stimuli. Both the voltage time-points from EEG and the independent component (IC) transform activations were normalised (see Fig 2.4). The kernel used with the SVM was a radial basis function (Keerthi and Lin, 2003) from the Matlab implementation of libSVM (Chang and Lin, 2011, Software available at the libSVM website).

### Matlab code change for negative control

```
ev_times_rand1 = randperm(total_time_ind,training_n);

if neg_control==1
    ev_times = ev_times_rand1;
end
```

Figure 2.2: For the negative control, the 'event times' vector is set to a short random permutation of the time index. This is set for each block. Thus, training and test data are deliberately misaligned.

## Classification results from the negative control

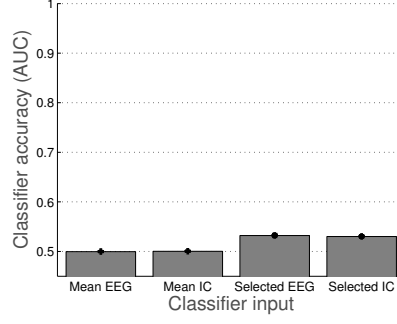


Figure 2.3: When using randomly misaligned trial labels for training, the accuracy of our classification algorithms is close to 0.5 AUC, which is chance classification. This is true for EEG and IC input. Standard deviation is shown in small black ticks.

The radial-basis function (RBF) kernel was chosen for faster training time in this implementation (Chang and Lin, 2011). We chose to use a RBF SVM as this can handle a wide variety of use cases and has fewer hyperparameters to set than using a polynomial kernel. This left  $C$  and  $\gamma$ , the two RBF hyperparameters to be set. In order to set a reasonable value for these without overfitting or exposing too much data, we performed a hyperparameter grid search – varying each of these – on a small subset of data.

To find appropriate values for the SVM hard margin training error cost ‘ $C$ ’ and the radial basis function kernel parameter ‘ $\gamma$ ’, we used the libSVM parameter sweep tool. A grid search was performed on 18 parameter values between  $C = [10^{-2} \text{ to } 10^{10}]$  and  $\gamma = [10^{-4} \text{ to } 1]$  on data from two participants. This suggested values of  $C = 1$  and  $\gamma = 1/\text{number features}$ , and those parameter values were used for all models.

### Classifier input

Each object had five presentation trials. Classifiers were trained using four of these and tested using their accuracy at predicting object onset on the fifth trial.

The training data was EEG data beginning at the initial time of object presentation until some time after (initially 500 ms). For each object, we take data that contains that object as being as ‘positive’ training data, and label it appropriately. The additional comparable data that did not contain the image was termed ‘negative’ training data. These training data examples were labelled as ‘1’ and ‘0’ respectively when used with our classifier.

For each object classifier, positive ‘object present’ input data came from the 0.5 s after each presentation of that object. As we used four trials, with each trial giving 0.5 s of objects onset data, we acquired 2 s of image-present data for each object. As a sample rate of 256 Hz was used, this gave a positive example training vector of  $512 \text{ datapoints} \times \text{number of channels}$ .

Training also included negative examples, from EEG recorded while no stimuli

was shown. Negative ‘object absent’ input data came from 19.5 s of randomly selected intertrial data. Classifiers were trained using a proportion of approximately 10% ‘object onset present’ data and 90% ‘object absent’ data – 2 s of data to 19.5 s of data. Test data also used this approximate proportion of data.

Test data came from the one presentation trial of each object that was not used for training (rotated as explained below) along with further negative examples of random intertrial data in which no object was presented.

### **Cross-validation**

To improve both the robustness of the classifiers and their ability to generalise to new data, 5-fold (leave-one-out) cross-validation was used (Efron and Gong, 1983). This process reduces the likelihood of erroneous results, as multiple splits of the data are considered. Cross-validation was performed on data from each object by dividing the data into five splits, with a single object presentation trial in each split. We then iterated through five separate SVM models – each training on four of five trials. The remaining fifth trial was then used as a blind test. As we iterated through the cross-validation, each trial was used once as a test data.

Classification test results then came from the accuracy on classifying the respective unseen fifth, and the score was averaged across the five splits. All classification results reported are from this average of all five cross-validation splits.

### **Performance metrics**

The success of a classifier can be given simply as percentage correctly classified. This can be valid in many contexts, but does not clearly show that performance depends on both ‘sensitivity’ (true positives) and ‘specificity’ (true negatives). A receiver-operator characteristic (ROC) plot (Mason and Graham, 2002; Hand, 2009) illustrates both sensitivity and specificity – with the area under the curve (AUC) of the ROC of 0.5 signifying random chance prediction and 1 being perfect prediction. This was relevant here as over 90% of data in both training and test situations belonged to the ‘object absent’ class. If a classifier were to predict ‘object absent’ everywhere, it might get 90% accuracy despite conveying no useful information. AUC, however, would correctly score that classification as no-better-than-chance performance.

Plotting a ROC curve can be particularly useful when sensitivity and specificity are being manipulated separately, but here we simply use area under this curve (AUC) as a concise metric of both classifier sensitivity and specificity.

#### **2.0.5 A note on chance and baselines**

For a ‘coin-toss’ classification of two equally probable outcomes, chance performance from random guessing would, of course, be 50%. Sampling from a sufficient number of trials – perhaps even 20 – would reveal this 50% accuracy clearly and simply. Area under

the receiver-operator-characteristics curve (AUC), where misclassifications reduce the score, would be 0.5 AUC.

The performance of a random guesser can also be clearly assessed in the slightly more complex situation where the probability of the outcomes is unbalanced. Taking two outcomes to be either rolling a 6 on a standard six-sided die, or rolling a 'not-6', always guessing the latter category (the majority class) would be correct in 5/6 (83%) rolls of the dice. Even with 83% 'correct', this is clearly not revealing useful classifications or predictions. Using AUC here quickly and concisely shows this to be a useless classifier, at 0.5 AUC.

In problems like the coin-toss, the 'chance' baseline classification rate is clear. Where the training and test data contain outcomes that are not equally likely, using AUC makes performance more easily apparent.

For situations in more complicated domains, with more involved data processing, classification tools and analyses, AUC can still give a good indication of performance. A problem with more complex situations is that there are more possible points of failure. Especially when many samples and classifiers are being used, it is important to take account of these multiple comparisons, and apply correction wisely. Classification performance, even with high AUC, should be considered in the light of what AUC might emerge by chance under similar conditions.

With this in mind, I designed a 'negative control' classification test to give a staunchly empirical indication of the random chance baseline classification rate. We use the exact same experimental data, pre-processing, data partitioning, classification code, stats and analysis methods as in the main results, but add a random offset to the trial start time. This means the training and test data that is grabbed is scrambled and so any trial classification is spurious – or at least not related to trial variation.

To some extent, the 5-fold cross validation method we use provides some resistance to spurious correlations, as does the blinding of test code and test data. The additional benefit of this negative control is to give an empirical and quantitative indication of the random chance baseline classification rate using this data pipeline.

Running our processing and classification algorithms on randomly sorted trial data gives low accuracy and classification performance (see Fig 2.3). For almost all iterations of classifying from the different EEG channels and IC data sources, performance is close to 0.5 AUC, with mean accuracy of 0.5003 AUC. This is consistent with chance classification, as expected. Using the channel selection procedure, we only select the best of a bad bunch, and so do no better than 0.53 AUC.

Using this negative control, we see that chance AUC is indeed 0.5 for average input, and channel selection is 0.53 AUC. As this was consistent across many runs, AUC much larger than this is indicative of a non-spurious relationship.

In order to quantify this in more detail, we can run Wilcoxon signed-rank statistical tests on experimental results vs results from the negative control. Running this on our main results, we find  $p \ll 0.01$  of selected EEG, cleaned EEG, or selected IC to

equivalents in the negative control classification. For point of reference, an AUC of above 0.6 will likely be significantly above chance in this protocol.

## **2.1 Experimental Methods 1 - Visual Classification I**

### **2.1.1 Participants**

Seven participants (5 women, 2 men; median age of 25) took part in the study and were each compensated £7 per hour. Ethical approval was granted by the University of Edinburgh Psychology Research Ethics Committee. Participants self-reported normal or corrected-to-normal vision and gave informed written consent.

### **2.1.2 Materials and design**

Participants were instructed to observe a series of object images presented to them on a computer screen. Fifty colour photographs of common real-world objects were selected from the Bank of Standardized Stimuli (BOSS) provided by (Brodeur et al., 2010). Multiple presentation trials of each object were required for classifier training and testing. There were five blocks of 50 trials, where each object was shown once in a given block, in random order.

### **2.1.3 Apparatus**

The experimental script was written in MATLAB 2009b (Mathworks, Inc., Natick, MA, USA). Extensions from Psychophysics Toolbox 3 were used for better timing precision (Brainard, 1997; Kleiner et al., 2007). Stimuli were displayed on a fast-refresh 22 inch Samsung SM2233RZ monitor at 1000×1000 pixels centred on a 1680×1050 pixel display, with the participants leaning on a chin-rest 72 cm away, giving an object size of 15°–22° in the visual field. The complete screen was around 38° by 24°.

### **2.1.4 Experimental procedure**

Each trial began with a central fixation cross for 300 ms, followed by a randomly selected object image that was presented for 1.5 s (Fig 3.1). An on-screen text prompt then asked the participant if they had recognised this object as something familiar to them. This was implemented to avoid passive viewing of the stimuli. Participants responded using buttons on a gamepad – one marked ‘yes’ if this object was recognised as familiar and one marked ‘no’ otherwise. The text was replaced by a blank screen on button response or after 1.5 s. There was then a random intertrial interval of 0–1.5 s, where the screen remained blank until the next trial began. The experiment was performed in a single session per participant of around 1.5 hrs, in which stimuli presentation and breaks lasted around 25 minutes.



## **Eye tracking via eyelink**

For some experiments, an SR Research EyeLink 1000 with 2000 Hz upgrade was used to track eye movement. Eye tracking calibration and validation was performed before testing and at the beginning of each experimental block (roughly 20 minutes). This was connected to a dedicated PC running SR Research EyeLink OS and acquisition software, which was connected over ethernet to the presentation PC that controlled the eye tracker using Matlab commands via Psychtoolbox (Kleiner et al., 2007). Eyetracking data was examined, but not used for classification.

### **2.1.5 EEG recording**

EEG was recorded from 64 head electrodes and 6 support electrodes using a Biosemi ActiveTwo amplifier at a sample rate of 1024 Hz. These 64 electrodes were placed according to the standard 10/20 EEG electrode system, and held in place using a Biosemi electrode cap of appropriate size for the participant. Triggers generated from the experimental code were recorded on the EEG device to allow timing synchronisation. All electrode offsets were below 20 mV. The six support electrodes were placed on 2 mastoids, 2 temples, and above and below the right eye, for better EOG detection. The experiment took place in a shielded experimental room.

## **EEG data processing and ICA**

This processing of the raw EEG data was performed using custom code that included use of standard EEG processing functions from the EEGLAB v12 toolbox (Delorme and Makeig, 2004) in MATLAB. Biosemi data was loaded using left mastoid reference, and re-referenced to an average reference later (Luck, 2005). A Hamming-windowed FIR band-pass filter of 0.1 – 80 Hz was applied, using ‘eegfiltnew’ in EEGLAB v12. Some trials with extreme values were automatically tagged using EEGLAB functions. Those electrodes with many extreme values were rejected (11-22, median of 20 of 69 removed). Independent Component Analysis (ICA) was applied to the whole dataset from each experimental session, after artefact channel rejection, using the Infomax ICA algorithm (Delorme et al., 2007). As 69 EEG channels were recorded, and around 20 removed for excessive noise, standard Infomax ICA gave a median of 49 ICs for each participant. The timing triggers were used to label the time-points in which each image was presented.

## **ERP waveforms**

After data processing, we confirmed that our visual stimuli presentation elicited a visually evoked potential change. EEG trace data was visually inspected (Fig 2.5) to check recording quality for EEG artefacts. Fig 2.5 (top) shows EEG trace from one trial on 10 EEG channels. No obvious EEG artefacts are present. The thin blue line

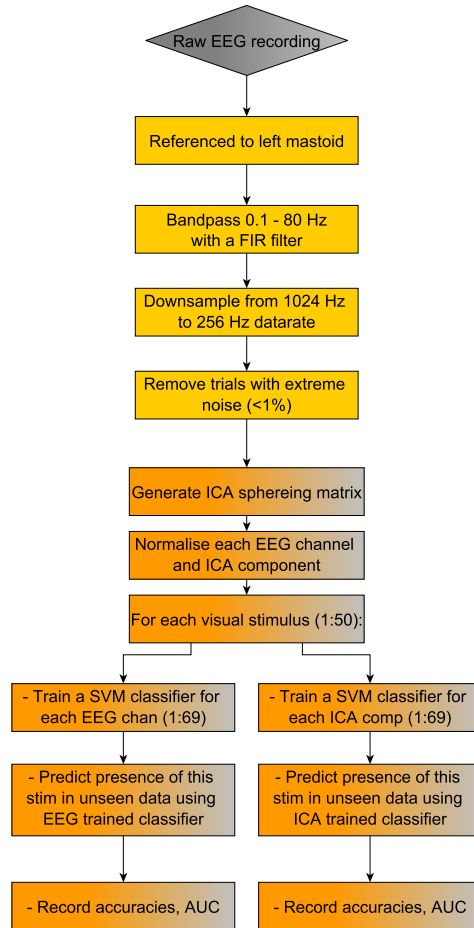


Figure 2.4: Data flow for EEG processing. Data processing steps are shown in yellow and machine learning steps in orange. The input here was the scalp EEG recording, which was referenced to the ‘mastoid’ electrodes behind the left ear. A FIR filter was used to reduce signal outside 0.1 Hz to 80 Hz. The sample rate was then reduced to from 1024 Hz to 256 Hz to speed subsequent analysis, and electrodes with very high electrodes noise were rejected. The ICA was run and SVM classifiers were trained for all objects using EEG channels data (left) and IC activation data (right).

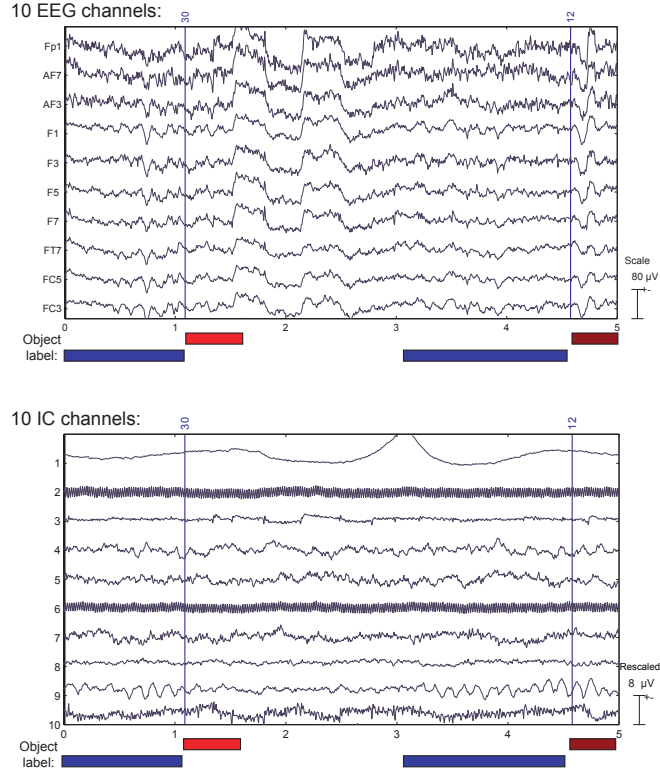


Figure 2.5: EEG trace and IC activations in an example trial. The top panel displays example EEG trace data from 10 channels for 5 s around the first visual object presentation (participant 1, trial 1). The lower panel shows the data from the same period in 10 channels of ICA-generated IC activations. The blue vertical lines show object presentation time (object 30, ‘bottle’, shown at around 1 s and object 12, ‘asparagus’ shown at 4.6 s) and the red bars indicates the approximate time labelled as ‘object present’ used for training classifiers. The intertrial time (blue) was labelled as ‘no object presentation’.

around 1.1 s and 4.6 s into the recording indicates presentation times of two objects, with timecourse labelled below in red.

## 2.2 Experimental Methods 2 - Online Analysis

From previous experiments, we have a machine learning paradigm where we detect presence of visual object given EEG data of that trial.

Several studies suggest that there are properties of ongoing EEG that are related to visual perception (Mathewson et al., 2009; Weber and Thompson-Schill, 2010; Thut et al., 2012).

As we were getting around 90% accuracy at labelling the visual state using only data from one IC, a pertinent question is 'what is it that the brain is doing that is generating this EEG signal?'. Is it the case that the differential response of the early visual system is generating this activity?

It is known from both large-scale anatomy (Mountcastle, 1997) and functional studies (Kay et al., 2008) that there are many, many recurrent connections between higher visual areas and V1, and from V1 to thalamus. It has been suggested that directed attention may be mediated through these recurrent feedback connections to lower-and-mid visual areas (Zhaoping, 2014). If we can profile a high-predictive IC source matching this in occipital cortex, can we probe whether this is under conscious control of the subject? How tightly does this fit to conscious perception of attentional focus?

In an experiment where we expose the subject to two visually prominent objects overlaid, and instruct them to bring their attention to one of them, can we use the activity of this IC to identify which object it is that the subject is attending?

Which ICs are more strongly predictive of stimuli or attention? At what timescales?

We suggest that our EEG classification protocol is a strong tool investigating such aspects, and describe one such experiment here.

### 2.2.1 Use of the online attention interface

An ambitious – but potentially revealing – extension to classifying vision from EEG is to investigate conscious modulation of the neural activity. That is, can a subject voluntarily choose to control the activity of the IC sources?

This is an important question, as this could give information as to the nature of the brain that would be difficult to acquire through other means. Empirical evidence from this could constrain the network-level properties of cognitive models, in a way that cellular or fMRI investigation would not be able to reveal.

As discussed in the Online Analysis introduction (see section 1.5), offline serial image presentation trials do not give data amenable to investigating attentional modulation.

Cerf et al. (2010) address this with the use of a novel design. Images are still presented in each trial, but there is the addition of a second image, overlapping with the first. With instruction to focus attention on one of these images, one image can be considered the 'target' and one image the 'distractor' (see Fig 2.6).

As each trial is ongoing, classifiers trained using neural data are classifying the live neural data from this trial. If the current neural data is closer to that in the attentional

target image, a step towards trial 'success' is recorded, or a step towards 'failure' for non-target output. Over a 10s trial, there may be some movement in each direction, but the net movement gives an indication of how well neural activity can be classified as attentional target.

These steps are indicated with feedback to the subject in real time. In this way, the subject gets feedback of the attention classifier as the trial is ongoing. Investigating the level of conscious control that a subject has over this activity is of central interest.

## Cerf duelling images online attention interface

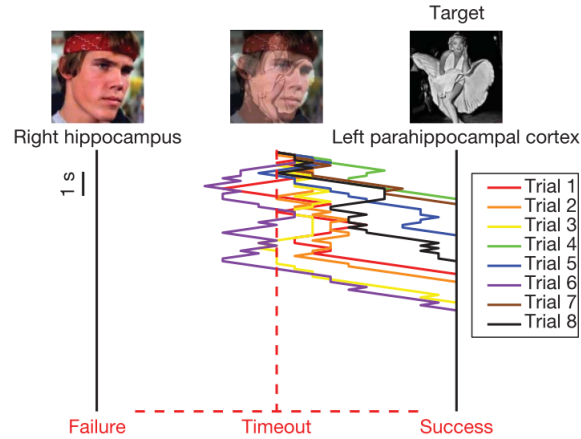


Figure 2.6: In Cerf et al. (2010), subjects observe overlapping stimulus of two faces, with the instruction to focus on the 'target'. Neural response is classified, and neural response matching 'target' classification yields a 'success'. On each 'success', the line graph shows progress towards the rightmost bar.

This can only be investigated when the subject has instantaneous feedback of the outcome of the attentional classifiers. Thus, the online attention interface is required. The success/failure line chart (as shown in 2.6) is an efficient way to relate this information, and so this was adapted for our experiment here.

### 2.2.2 Motivation

In Chapter 2, we developed a system of using machine learning on EEG recorded from participants in a visual object presentation experiment and found that this could be used to read the identity of the presented object from around 0.5s of EEG signal at above chance accuracy.

We wish to investigate the process of human visual perception, and suggest that this system could be used for novel experiments in that area.

The accuracy of classifiers trained on relevant data might be thought of as a proxy for information content within a specific subset of the EEG recording. If this is done while the subject is in different state of visual perception (perhaps beginning with perceiving 'apple'; perceiving 'banana'; perceiving neither), then we can assess possible information content predictive of each of those states from different data sources of EEG

electrodes, ICA sources, temporal latencies, frequencies, and other data transforms.

There is some evidence that it may be useful to consider active visual attention, passive viewing, and visual perception as distinct processes - both conceptually (Zhaoping, 2014) and as is implemented at the level of neural coding (Scalaidhe et al., 1999).

Can these distinct processes be addressed experimentally using EEG and machine learning? While experiments explicitly addressing the computational neuroscience of vision are often performed using lab animals, EEG has the advantage of acquiring data from awake human subjects. This means the subjects can report the subjective content of their perception, and also (frequently) follow instructions to attend to specific objects as part of the experiment.

So, given that a system of high quality EEG+ML may be useful in investigating information content flow in visual attention and perception, can we directly use this experiment and data from the Visual Classification I experiment to investigate this?

Unfortunately not. While the Visual Classification I experiment demonstrates classification of EEG data from subjects while they view objects, we suggest that in order to sensibly and selectively probe visual attention, there are several prerequisite conditions that need to be fulfilled:

1. data recorded from trials in which the subjects are shown multiple objects, and their attention is directed to one of these objects
2. confirmation from the subject that that object was indeed the target of their attention
3. data from the same object shown, but with attention not on this object
4. data from the same object, shown alone

With prerequisite 1, we can see that investigating attention might well require experiments with explicit competition for attention. Prerequisite 2 is required as, although subjects might well be instructed to pay attention to a specific stimulus, it may be that they fail to do so, through fatigue, weariness after many trials, or insufficient haste. Thus, a subjective report of what it was that the subject was attending to in each trial becomes useful. A trial in which the subject was directed to attend to the 'apple' should not be taken as an exemplar of a trial in which the 'apple' object was attended to without subjective verification.

Prerequisites 3 and 4 provide data that selective visual attention of this object can be compared to, and perhaps contrasted against passive viewing or observation of this object as a distractor.

### 2.2.3 Designing the online experiment

The earlier Visual Classification I experiment was not designed to address visual attention and does not meet the above prerequisites. The subjects there did view objects,

but only a single object in each trial. Addressing visual attention and meeting these requirements necessitates a new experimental design.

In the design of a new experiment, are there other lessons from the previous experiment that can be included?

A possible modification might be to take account of the variation of each subject to each object. As shown in Fig 3.10, some displayed objects were detected at higher accuracy. This was also the case within subjects, with some objects in most subjects having substantially higher accuracy than mean. While there are many possible reasons for this difference (natural variation, personal salience of each image to each subject, chance, etc), a practical outcome is that most trials are from presenting stimuli which elicit a response that is not decoded as well as the highest accuracy decoded images. As investigating visual attention will likely give a signal that is more subtle, it may make sense to focus on stimuli that are more likely to elicit a larger detectable change in each subject.

Another consideration is the possibility of providing feedback. If models are trained to classify the selective attention on an object, assessing the degree of conscious control of this response could be uniquely informative of the nature of visual perceptual processing. For instance, imagine a classifier gives above-chance accuracy at decoding 'attention on apple' state with data from IC1. If the subject is shown the output of this classifier in real time, perhaps classifying the last 0.5s of data, can she consciously choose to increase the positive response of this classifier by looking at the apple? By focussing on the apple? Or imagining an apple, even? What is the latency of this response? Is there temporal decay of this, responding only at attention onset? Which IC sources are informative, and when? This can build on other work assessing voluntary control in vision, such as that in recording from electrodes in the temporal lobe of epilepsy patients (Cerf et al., 2010). With the presence of object-attention classifiers, identifying what is and is not under conscious, voluntary control could offer evidence in favour of competing theories of human visual processing (Zhaoping, 2014).

Both these changes to the Visual Classification I protocol require introducing online analysis. That is, rather than EEG data from the experiment being written to disk during the experiment, and loaded, processed, and analysed at a later date, feedback and selecting stimuli require that processing of the EEG occurs as the experiment is ongoing.

In order to access EEG data from the Biosemi recording system in real time, the conventional recording software had to be bypassed. Custom data handling tools were developed in Matlab to interface with the Biosemi driver and so access data with very low latency.

#### **2.2.4 EEG recording environment**

A Biosemi ActiveTwo was used to record EEG. This was used with specific drivers, connected directly and only to a single laptop. This allowed connection to Christian

Kothe's 'Lab Streaming Layer' through which we can acquire data from the system in near real time with latencies of  $< 20ms$ . I have written a Matlab interface for this which allows the use of this data with the standard Matlab and PsychToolBox stimuli display system.

With this new recording environment, a single laptop could be used for recording EEG, presenting stimuli, running the classifiers and data processing. This replaces the previous separate stimuli machine and EEG recording machine.

Elements of the recording environment:

**Participant** The human participant, instructed to watch screen and respond

**Electrode cap** A cap fixing 32-72 active EEG recording electrodes correctly to the head of the participant

**Biosemi ActiveTwo AD box** The Biosemi EEG amplifier, which does analogue-to-digital conversion at 24 bit resolution and connects via fibre-optic and USB to the laptop

**Laptop** A single quad-core computer, on which display, recording, analysis and classification is run

**Lab Streaming Layer (LSL)** Interface software that deals with data coming from EEG and can be polled from Matlab with greatly reduced latency and data loss

**Matlab** High-level scientific programming language

**EEG analysis and classification code** Matlab code that we have developed to import EEG data quickly from LSL as a Matlab object, run EEG data processing on, build SVM models and classify data to best label it.

**Psychtoolbox (PTB)** A 'toolbox' of software that integrates with Matlab to use low-level driver functions that (among other things) allows OpenGL and much more reliable timings of what is drawn on the screen. Developed by Kleiner et al. (2007) and Brainard (1997). We have Matlab code that uses PTB to display our experiment as desired.

**Monitor** A 22" fast-refresh 120 Hz screen which displays stimuli to the participant.

### 2.2.5 Materials and design

Object stimuli and presentation stimuli training used were similar to that in 'Offline classification' 2.1.2. A selection of everyday images from BOSS (Brodeur et al., 2010) will be shown using Matlab and PTB.

First, we can present object stimuli trials as before. We can do data processing and train classifiers as the session is ongoing.



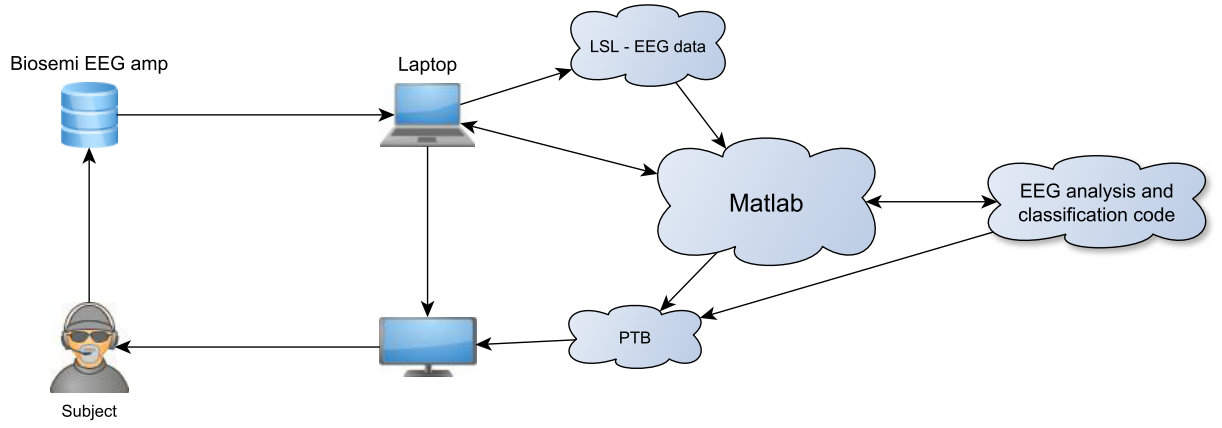


Figure 2.7: A schematic outline to illustrate dataflow in the online system. LSL refers to the 'Lab Streaming Layer' and PTB refers to our PsychToolBox presentation code.

With training data and classifiers ready, we can begin a visual focus competition experiment, as in Cerf et al. (2010). In each competition trial, two object stimuli can be presented, unsaturated and overlaid, and the participant is (before the trial) asked to 'focus' on a specific one. We can repeat many object competitions in many short (500 ms) trials.

On each object competition trial, if the classifier output is more indicative of the primed object, we provide a reinforcement 'beep' and update an on-screen graph to success.

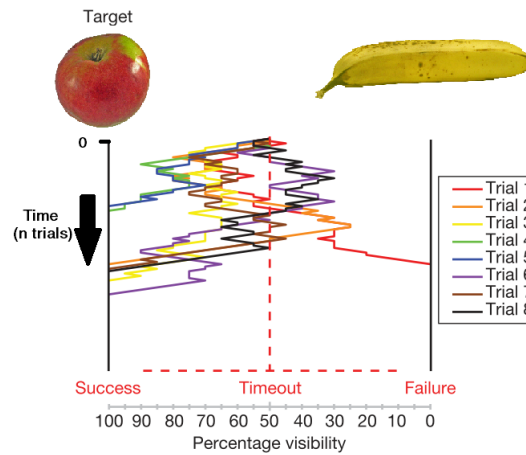







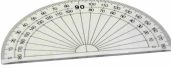


Figure 2.8: A chart of 8 example trials, with the competition judgement beginning at 0 and proceeding down in time and toward either success or failure after each presentation. If the presentation was judged by classifiers to be an apple, judgement takes one 'step' towards success. Trial ends when either success or failure is reached, or timeout after 10 seconds. Adapted from similar design in (Cerf et al., 2010)

Success at more than half of competition trials indicates that the classifiers identified target more. It could be that some ICs are able to classify object, some object and

Table 2.1: Example BOSS stimuli images used for Online Analysis Stage I

'log'	'pear01'	'globe'
		
'lifejacket'	'handfan01b'	'pitabread'
		
'mask02a'	'protractor'	'blank'
		

conscious target, and some ICs can classify neither. Can we identify these?

## 2.2.6 Experimental procedure

Initial trials will be object stimulus training, where participants report object recognition as in Visual Classification I. Blocks can be rearranged such that some objects are shown 4 times earlier on and SVMs can begin training models using this input straight away, as the experiment is ongoing.

The aim of this is to explore these signatures of visual objects in the EEG data that give higher classification performance and investigate these signatures when attentional focus is changed – not just stimuli. With this, perhaps we can figure out if we can use this to say anything useful about the computational neuroscience of visual perception and attention.

We can describe this experiment in three stages:

**Stage I** – Again, we repeat Visual Classification I, where the task is simply observing the images. We train SVMs on lightly processed EEG scalp and IC data as the stimuli is being displayed. Additional test data is recorded *after* models have already been trained, giving us more evidence of generalisability and consistency over time.

**Stage II** – Expanding on this, we then address the main question of this experiment – can we also separate stimuli attended to from stimuli simply present? For the task in this section, we direct the participant’s attention to 1 of 2 superimposed

objects. Are there consistent signatures of *conscious attention*, separate to those already detected for observation? What are their properties? When do they occur, relative to stimuli appearing? Where do they occur - in both electrode and IC source-space?

As these questions have not yet been fully investigated, we are still building our answer. To improve our chances, we can narrow the search to focus on those stimuli and data sources where we have good classifier performance in this participant, as is found on the previous step, and so this needs to be done online and within one recording session.


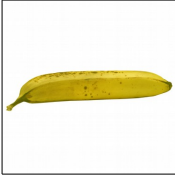

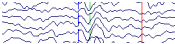
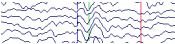
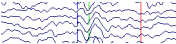
Once we have data from trials where we have been directing attention to certain objects (and recording subjective loci of attention) we can attempt to build SVM models where positive data is only those times where an object is present and participant reports attending to that object. The signal indicative of specifically attending a certain object might be quite small, so what kind of significant accuracy can we get? How different are the best attention-specific classifiers from the best object classifiers? What temporal overlap do they have? What spatial overlap? What can this tell us about attention and perception? Is this compatible with prevalent explanations in the literature?

**Stage III** – A third, more speculative, stage of this experiment then investigates the subsequent ability of the participant to *consciously modulate* the response of these signatures of attention. The task is now to bring attention to 1 of 2 objects while feedback is given of what object we think they are attending to most, given the current activity of attention classifiers.

The trials begin similar to those in stage 2, but the output of (for example) the apple-attention classifier is also displayed on the screen. Each 500ms, we poll the current output of the best attention classifiers when given the last 250-500ms of EEG data. If apple is the current target AND the apple attention classifier is outputting 1 AND the non-target attention classifier is outputting 0, a 'victory' beep is played, and a step towards 'victory' is shown on a simple plot. If the apple attention classifier is outputting 0 AND the non-target attention classifier is outputting 1, a 'loss' noise is played and we step towards 'loss'. If neither occurs, we advance one time step, moving the cursor down one step, but neither left nor right.

This section must be online for the feedback. It is also dependant on selecting models with higher accuracy from the previous stage. This is potentially incredibly informative, as we can assess if participants are able to intentionally change the activity of these parts of the EEG we are detecting, or if they are fixed in response to the stimulus.

**Stage I trial structure:**

	Trial 1	Trial 2	Trial 3	...
Stimuli displayed on screen				...
Class Label	Apple	Banana	Crayon	...
Recorded EEG				...

**Stage II trial structure:**

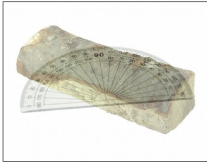
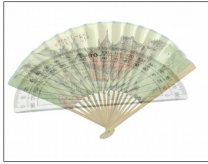
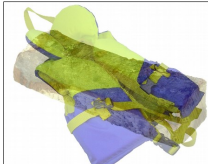

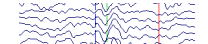
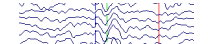
	Trial 1	Trial 2	Trial 3	...
Prompt	Please focus on the PROTRACTOR	Please focus on the FAN	Please focus on the LIFEJACKET	...
Hybrid stimuli shown on screen				...
Object class label	Protractor + log	Fan + protractor	Lifejacket + log	...
Attention class label	attn_protractor	attn_fan	attn_lifejacket	...
Recorded EEG				...

Figure 2.9: Trial structure of Stages I and II, along with example stimuli. We replicate stimuli presentation, get data on attention, and build classifier models on this data.

### Stage III trial structure:


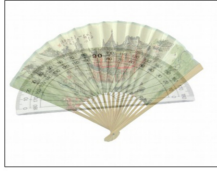
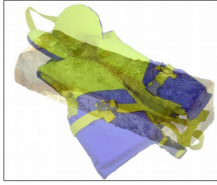
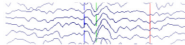
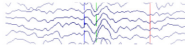
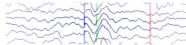
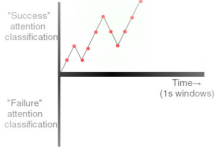
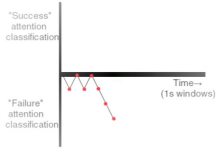
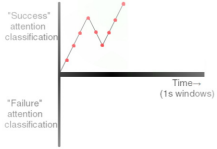
	Trial 1	Trial 2	Trial 3	...
Prompt	Please focus on the PROTRACTOR	Please focus on the FAN	Please focus on the LIFEJACKET	...
Hybrid stimuli shown on screen				...
Object class label	Protractor + log	Fan + protractor	Lifejacket + log	...
Attention class label	attn_protractor	attn_fan	attn_lifejacket	...
Recorded EEG				...
Attention classifier chart throughout trial				...

Figure 2.10: Trial structure of Stage III, along with example stimuli. Using models trained from Stages I and II, we classify attention focus in real time, giving feedback as each trial goes on.

### 2.2.7 Online analysis, Stage I differences from Visual Classification I

To investigate attention classification, we use the online attention interface. The experimental set up is much the same as in the main visual classification experiments specified above. The trial structure, example stimuli and the interface for the 3 stages can be seen in Fig. 2.9 on page 46. In Stage I, we gather data from single image presentation trials, presented alone, largely as in the main visual classification experiments. We use a reduced subset of 8 of the BOSS images (Brodeur et al., 2010) along with two 'dummy' blank images, presented in random order, 15 times each for a total of 150 trials. Stimuli were on screen for exactly 0.5s per trial.

In contrast to earlier recordings which wrote EEG data to be analysed offline, data was streamed from the Biosemi amplifier to Matlab environment using the Lab Streaming Layer tool. In stage I, data was polled for each trial, and saved in an array structure.

On the completion of the stage I stimulus presentation, models were trained to classify those trials. Electrode normalisation values were saved and the ICA mixing matrix was computed. For each subject, the ICA mixing matrix generated in stage I was retained for later stages for that subject, and not recomputed. To run the classification

analysis, the code from Visclassica was mostly reused, but amended to have a one-vs-rest structure, with a 'positive' label for each object in turn and 'negative' label for the others.

### 2.2.8 Neurofeedback setup for stage III

With stage III, we want to investigate the conscious control that the subject may have over the attention-specific neural activity. Can this be modulated at will?

On completion of stage II, attention-specific classifiers trained from stage II data, with 'positive' labels for attention on an object, and 'negative' labels for attention on another object.

For stage III, the stimuli was again two overlaid BOSS images. Within each 10s trial, data was continuously polled from the Biosemi EEG amplifier. The most recent 400ms of EEG data was transformed using the existing ICA mixing matrix. This data was then immediately used as a test example for the attention-specific classifiers, and the output was recorded. This was repeated every  $\sim 1$ s throughout each 10s trial.

In addition to the image stimulus, subjects could also see a small line graph showing the outcome of the attention classifier (as shown in Fig 2.11). Each time the attention classifier output matched the target, a 'ding' reward sound played and the line graph stepped toward 'success'. If the attention classifier output did not match, a 'buzzer' sound played and the line graph stepped down toward 'fail'.

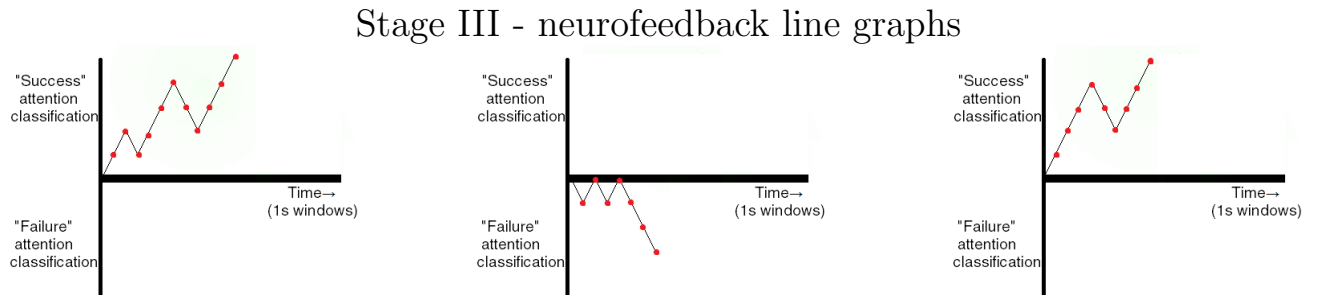


Figure 2.11: Three examples of the neurofeedback line graph. These display to the subject the ongoing success or failure of attention classification, matching the current attention target, updating roughly every 1s.

At the end of the trial, net 'success' or 'failure' is assessed by the net distance from the height of the line graph above the origin.

Only a few subjects reached the stage of the experiment where we could investigate this, due to technical problems with the display code and recording hardware.

### 2.2.9 Online system implementation

While online recording, processing, classification, and feedback were required for this experiment, having this system run online is quite different from the standard offline

EEG recording setup, and this had not previously been attempted within our PPLS EEG lab at Edinburgh.

An initial technical hurdle was how best to perform online data acquisition from our Biosemi ActiveTwo system such that this might be used with our existing setup. While the proprietary Biosemi acquisition software does allow some online analysis and customisation through use of LabVIEW (Biosemi, 2013), this was found to be somewhat inflexible for our purposes. An alternate data acquisition system was found in Christian Kothe’s BCILAB and LabStreamingLayer, developed at the Swartz Center for Computational Neuroscience at UCSD (Delorme et al., 2011). With this, we have specialised drivers for the Biosemi USB receiver, along with a separate dedicated C++ program that handles the Biosemi I/O, reading the incoming data samples. LabStreamingLayer creates a buffer, from which this data can be read in real time from within the Matlab environment. Thus, with our adaptations, the standard, well-tested EEG processing code from the EEGLAB Matlab toolbox could be utilised online.

It was crucial to address some setup practicalities of this novel online system in order to have it work well.

### Noise reduction

As EEG is dealing with small variations in the electric field, it is subject to the prevalent electric field noise in the environment. Standard practice in offline EEG analysis is to run batch trial averaging and filtering on each trial some time after the experiment has been run. As we wished to use EEG information in real time, it was not feasible to do post-hoc averaging and filtering. A reduced set of signal subsampling, filtering and noise rejection needed to run online.

With this, there is increased impetus to ensure that signal quality is high. With the Biosemi ActiveTwo system, the physical process of affixing electrodes (and ensuring that conductive gel makes good contact with the electrode and the participant’s scalp) is not entirely error-free, and so a small number (perhaps 3 of 64) might be ‘lost’ over the course of longer experimental session. In post-hoc analysis, these would be ‘rejected’, or deleted from the recording data (Luck, 2005).

With online analysis, there is room to improve this process. Rather than examining all data in a batch after the recording session, we can plot metrics of signal quality as the experiment is ongoing - and even pause the experiment if any electrode becomes loose or noisy (see Fig 2.13). In order to compare noise in different electrodes, participants, sessions, and experimental setups, a concise metric is useful. While there are many ways that ‘noise’ might be quantified, we used a simple invented heuristic of ‘EEG jitter noise’ - absolute change voltage over time, in each sample to the next, and summed over all samples in this chunk. This is shown in table 2.13, where  $m$  = length of chunk (typically 256 samples  $\approx 0.5s$ ) and  $sr$  = sample rate in Hz.

$$\text{EEG jitter noise} = \frac{sr}{m} \cdot \sum_{i=1}^{m-1} \text{abs}(\text{voltage}(i+1) - \text{voltage}(i)) \quad (2.1)$$

Intuitively, this EEG jitter noise can be thought of as change in voltage over each sample. This is scaled to a timespan of one second with the  $\frac{sr}{m}$  term, and has units of sample  $\Delta V/s$ , or typically  $\Delta\mu V/s$  here.

The online identification of 'noisy' channels allows those electrodes to be adjusted and perhaps fixed as the experiment is ongoing. As classification is performed online, there is also less opportunity for extensive offline batch data cleaning. This means there is perhaps increased impetus to ensure that electrode recording quality is high.

This was calibrated empirically and liberally, removing a few electrodes for each subject. A period of instructed heavy blinks was used to find possible values for upper bounds.

In Fig 2.13, we can see the online identification of 'noisy' EEG channels, using this method. In order to check if this noise identification technique is valid, several small tests were run.

Initially, we compare an EEG recording from a known good session to a known good session but with heavy blinks, and to the same session but before the electrodes have been secured and fully gelled.

## Exploration of electrode noise metrics

### 'Jitter noise' heuristic

Recording source	Jitter noise mean score	Standard deviation
S1 Known bad (before complete gelling)	$3.90 \times 10^4$	$8.00 \times 10^4$
S1 Known good	$1.35 \times 10^3$	$6.95 \times 10^2$
S1 Known good with heavy blinks	$1.40 \times 10^3$	$6.95 \times 10^2$
S2 Known good	$4.55 \times 10^3$	$1.6 \times 10^4$
S2 Known good with heavy blinks	$1.83 \times 10^4$	$1.58 \times 10^4$

Figure 2.12: Using the smaller 'jitter noise' heuristic, as specified in 2.1, we can see that mis-gelled electrodes do indeed have high 'jitter noise', instructed periods of heavy blinks are less bad, and known-good normal trial time is lower still. This is useful in determining what that noise threshold should be.

From exploratory use of this 'jitter noise' metric as shown in the electrode noise table, we can see that it is indeed sensitive to electrodes with poor connections, as when some electrodes are missing gel, the jitter noise score is an order of magnitude higher than known good recordings. In contrast, it is less sensitive to normal EEG variation on the scale of blinks, as there was little change in jitter noise scores in trials containing many blinks. In both participant 1 and participant 2, there was a moderate increase when blinking, and a larger increase with disconnected channels. There is some variation in the noise score across participants and sessions, but within a session it is quite stable.



From this, we can see that this jitter noise might be used to record large changes in electrode recording quality. If an electrode exceeds a participant-specific threshold (which will likely be on the around  $2 \times 10^4$ ), then this can be flagged in real time and corrected.

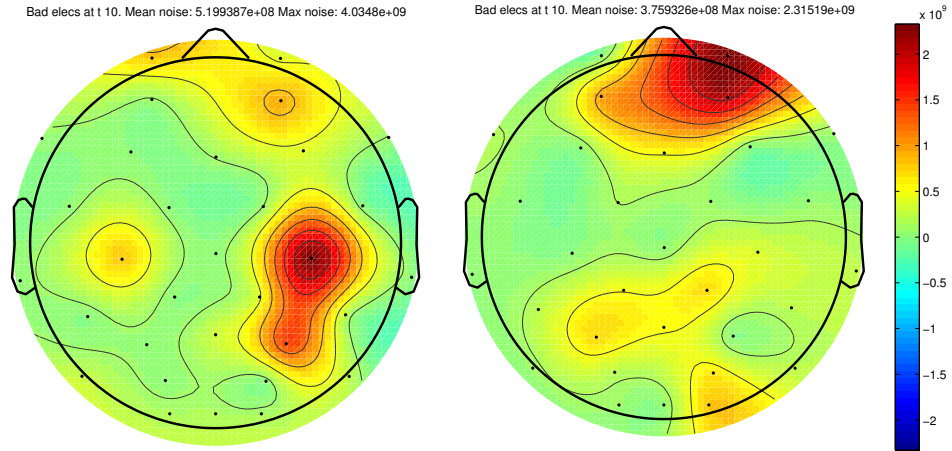


Figure 2.13: More exploration of possible noise heuristics. A metric of jitter noise is plotted on a headmap. Online tracking and visualisation of electrode noise, in sample  $\mu V$  change per second. Electrodes (here, 32) are plotted on a 2D headmap, with the nose shown at the top. Here, the participant on the left has one central-right electrode with noise of around  $4 \times 10^9$ . On the right, two frontal-right electrodes (Fp2 and AF4) are at higher noise levels. In each situation, these electrodes can be re-applied with gel, improving signal quality.

### 2.2.10 Advantages of online

It appeared that there were several advantages to this adapted EEG recording environment. With data recorded and analysed in real time, we could observe single-trial ERP as the experiment was ongoing. For example, through initial use of the online experimental environment in combination with the jitter noise visualisation, it was possible to quickly and readily identify sources of noise contamination more easily than through the standard Biosemi Actiview software.

Even when already following best practices for recording labs (Luck, 2005), through this methodology we were able to identify errant sources of noise. We found that a specific power extension cable and computer speakers left in the experimental lab greatly increased noise level. There was also a very large increase in signal noise when the laptop power charging cable was connected, although most recording had been done on battery power. The CRT appeared to increase noise slightly, the fast Samsung LCD less so, and the battery-powered laptop screen even less so (see appendix section A.1 for monitor details).

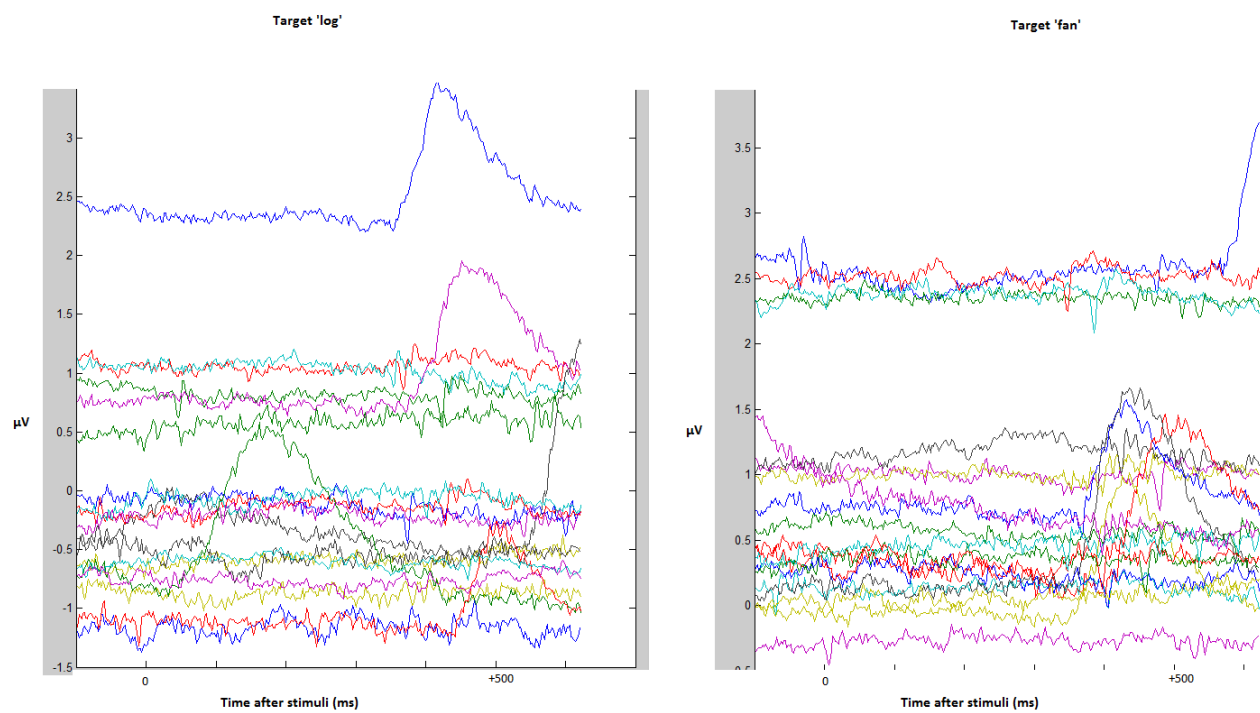


Figure 2.14: Adjusted raw (unfiltered, no baseline) ERP responses from electrodes C2 in participant 1. As single trials were run, data was displayed here. Each line shows a single trial.

## Chapter 3

# Results

### 3.1 Results of Visual Classification I

In order to investigate the application of machine learning in profiling EEG data in visual object presentations, we first ran a simple experiment to acquire data. Fifty objects were repeatedly presented to participants and this data was analysed offline.

#### 3.1.1 Hypotheses tested here

We proposed a method for classifying visual object presentation from EEG using machine learning. With this experiment, we hypothesised that in using SVMs, we can classify the presence of visual object stimuli at above chance levels, addressing Core Hypothesis 1 1.6.

Further, the use of ICA data processing might act as a feature selection step, improving accuracy and providing possible EEG ‘sources’, addressing Core Hypotheses 2 and 3 1.6.

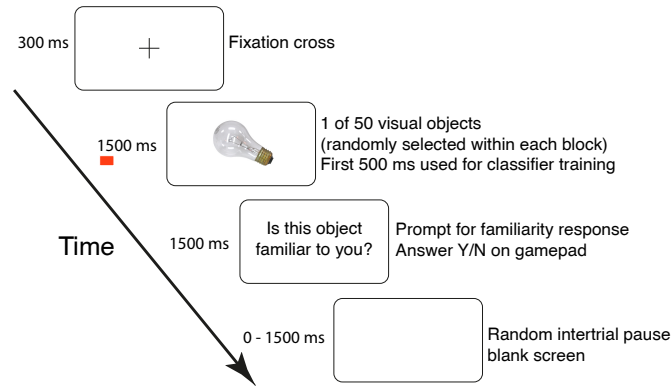


Figure 3.1: Experimental design of the visual object presentation protocol. The time-course of the events is shown. Participants were shown a fixation cross before stimulus of an everyday visual object – in this example, a light-bulb – compiled from a published standardised image bank (Brodeur et al. 2010). There were 250 of these trials – five presentations each of the 50 stimuli, ordered randomly. The red bar indicates the 500 ms after presentation that was used as object onset data to train classifiers.

### 3.1.2 Classification of visual stimuli

We found that classification of visual stimuli state in single-trials was possible, first using 500ms of EEG data per trial with these SVM classifiers. The data was labelled with either ‘object present’ (first 500 ms after stimuli presentation) or ‘object absent’, and the classifier was to label subsequent data appropriately, with a higher AUC indicating more trials correctly labelled.

Using all 49 channels of EEG data gave slightly above-chance trial classification performance. In every participant, we found increased accuracy using a single channel, rather than training models using all data channels simultaneously. Classification performance was improved when using one channel of EEG data, rather than all 49 channels at once. Average accuracy was increased from around 0.51 AUC to 0.65 AUC. Further, we found a small subset of channels that consistently gave much higher classification accuracy. To identify these, we used an automated selection procedure, based on data separate from the test data.

Specifically, data from 10 of the 50 objects was separated. On this data, classifiers were trained using data from each individual channel. That is, 49 separate classifiers were trained for each object, using data from each of the 49 data channels. The single data channel that gave highest AUC on these first 10 objects was then used as the single ‘selected’ channel. This data selection procedure was performed once on each participant for both EEG and IC data (see below). The selected channel was used as the best input to classifiers for training SVM on the remaining 40 objects. In the case of using scalp EEG data, this selection of one channel input increased classifier performance from 0.53 AUC mean to 0.65 AUC (Fig 3.3).

As shown in Fig 3.3, test data was correctly labelled well above chance performance

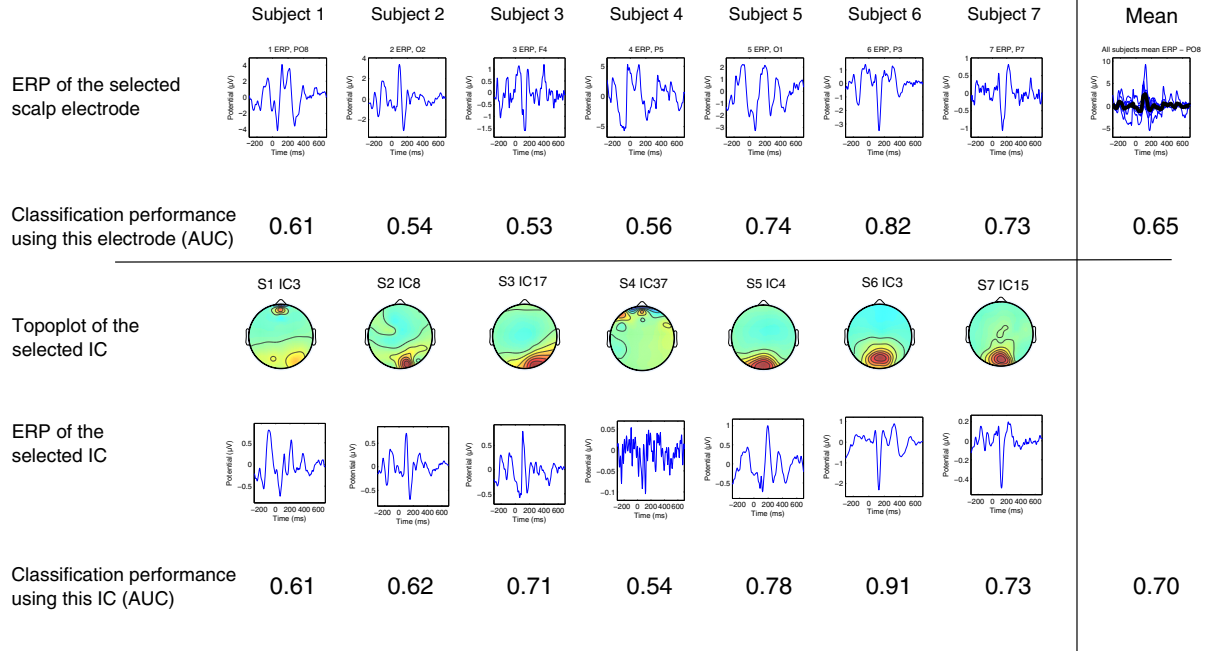


Figure 3.2: Top – Scalp EEG ERPs from the selected electrodes in each of the seven participants, along with the object classification performance when using data from that electrode. Thus, electrode PO8 is the selected electrode in participant 1, and when data from PO8 is used as input for SVM classifiers, object presence is correctly identified at 0.61 AUC. In participant 6, P3 is found to be the better electrode and it gives 0.82 AUC in that participant. Bottom – In the lower section, we show details of the topographic plots of the Independent Components identified by ICA, using EEGLAB ‘topoplot’. The selected IC for each participant is shown. As these ICs do not have easy standardised names to refer to (as EEG electrodes do), we show the spatial composition of these ICs on a 2D headmap, as if one were looking down on the scalp with the participants’ nose at the top of the topoplot. We also show the IC ERPs and classification accuracy using data from this IC. Note that mean classification accuracy is higher when using selected IC data than scalp ERP data, and that all those classifier inputs that give AUC of above 0.7 are located in the occipital-parietal regions.

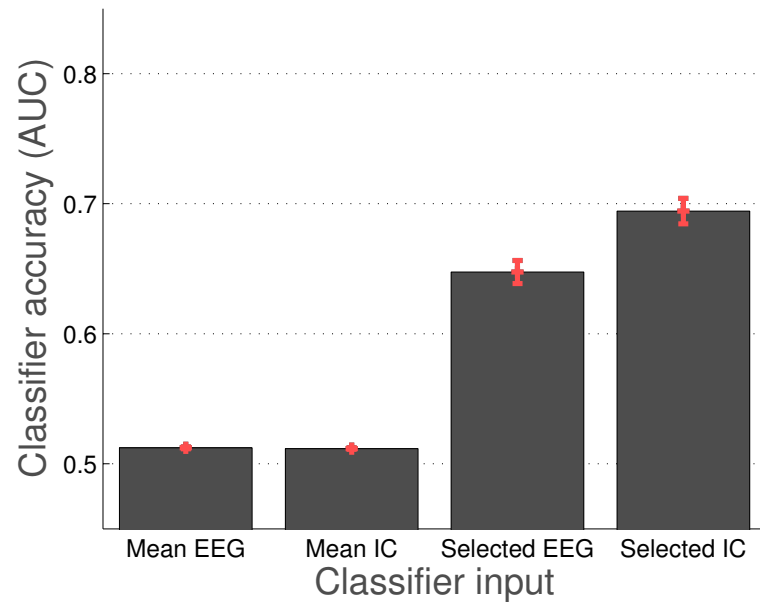


Figure 3.3: Error-bar plot of mean classifier performance across all participants and objects using EEG (mean of all channels), ICA (mean of all channels) or the selected channel for EEG and ICA.

of 0.5. This was the case when using input from a single selected EEG channel (column 3) or a single selected IC activation channel (column 4). However, we found accuracy on most channels was only slightly above chance (columns 1 and 2).

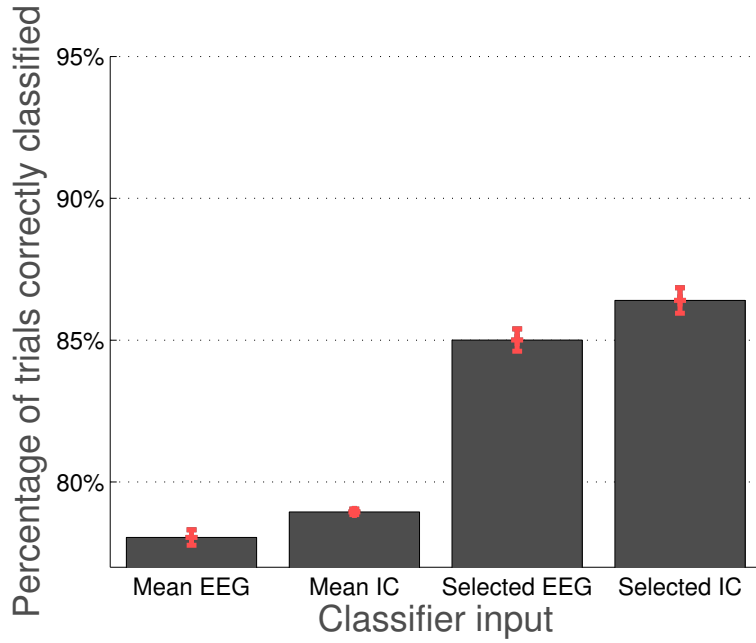


Figure 3.4: Error-bar plot showing mean classifier performance in terms of percentage of single-trials, where adjusted chance here is 80%.

Table 1: Accuracy of visual object classification (AUC) on each participant, using selected classifier input data

Input	EEG Selection (AUC)	IC Selection
Subject 1	0.61	0.61
Subject 2	0.54	0.62
Subject 3	0.53	0.71
Subject 4	0.56	0.54
Subject 5	0.74	0.78
Subject 6	0.82	0.91
Subject 7	0.73	0.68
Mean	0.65	0.70

### 3.1.3 Classification using ICA data

We found that training using data from a single selected IC increased average classifier performance from 0.52 AUC to 0.70 AUC (Fig 3.3).

Using a 1-tailed Wilcoxon ranked sum test (a paired MannWhitney U), classification accuracy was shown to be significantly higher when using selected ICA channel than using a selected scalp EEG channel ( $p < 0.001$ , means 0.70 AUC and 0.64 AUC).

Average classification accuracy was further improved when using data from a single selected channel of IC activation. This IC activation data was acquired by running ICA on each participant’s entire EEG recording. The same single data selection procedure

was used in ICA as was used in EEG. This gave an average of 0.70 AUC (Table 1), with around 87% of timepoints being labelled with the correct visual object presentation state, which was an increase of more than 4% above the accuracy on using the single selected EEG channel.

These data suggest that EEG channels and IC activations differed in their predictive power. This is further supported by participant-based heat maps (see Fig 3.8), which displays the accuracy using every channel of EEG or ICA data on classifying each of the 50 objects presented. Specifically, the plots show classifier accuracy across the all EEG channels (left panels) or all IC activations (right panels). In these maps, each column represents an EEG channel (left) or IC activation (right). The heat-map columns are arranged by AUC, from highest to lowest data input column. This indicates that the AUC from an average EEG or IC input channel are similar, yet the best IC outperforms the best EEG channel.

#### **3.1.4 Percentage correctly classified**

These results could also be summarised in terms of mean percentage of trials correctly classified. We show this in Fig 3.4. We find these agree with the AUC reported above in Fig 3.3, with the selected IC giving highest average classifier performance of 87% of all trials given the correct of object presence.

#### **3.1.5 EEG ERP and IC details**

Mean scalp EEG ERP is plotted in Fig 3.2. On presentation of visual object stimuli, we found the mean grand-average ERP with a prominent positive voltage deflection around 100-200 ms after presentation on central and parietal electrodes. This is in agreement with ERPs observed elsewhere – e.g. visual object presentation in Rousselet et al. (2007). Component activations over time from Independent Component Analysis (ICs) generated from this same data were plotted for comparison (Fig 3.2 lower). This trace data (Fig 2.5) and averaged ERP data (Fig 3.2) both appear similar to data recorded in other visual object presentation paradigms (Wang and Jung, 2012).

#### **3.1.6 Scalp EEG data with IC artefacts removed**

The above results suggest improved classification when using data from activations of components identified by ICA with these classifiers rather than using scalp EEG data, at least when a selected higher accuracy data channel is selected. A possible explanation for this might be that relevant neural signals are overwhelmed by electrical artefacts that are not relevant to the visual state and these are present to some extent on all electrodes. Then, when ICA is run, the contributions of those artefactual signals are 'captured' in single ICs.

In order to assess this, we now remove the contribution of each of those probable artefactual ICs from the scalp EEG data and test the visual state prediction again.



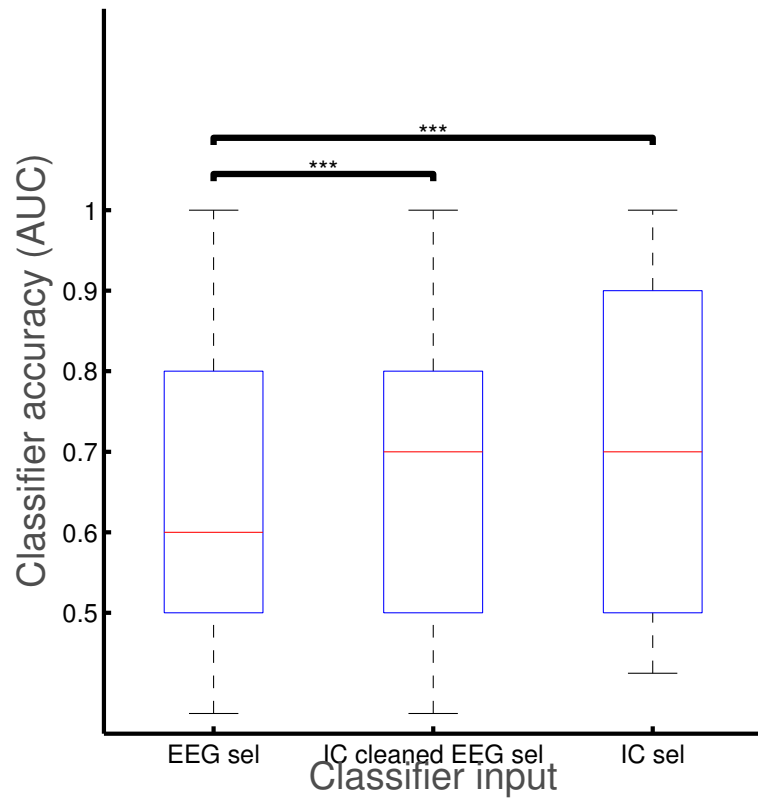


Figure 3.5: Comparing trial classification performance using EEG data, EEG data in which noisy ICs have been removed, and IC data. An average of 4.5 out of around 49 total ICs were removed. \*\*\* -  $p < 0.0001$

This effectively removes the contribution of these ICs from the EEG data.

The 'autorej' function of EEGLAB v12 was used to identify ICs with extreme values. Each component with activity above a rejection threshold of values of above or below 50 standard deviations was tagged as a possible artefact and the contribution it gave to that participant's EEG was removed.

This resulted in a mean of 4.5 ICs being removed (range of 1 to 12). Visual classification using this IC-cleaned scalp EEG data is shown in Fig 3.5.

Using ICA to remove the contribution of obvious artefacts from the scalp EEG data improved the accuracy on this 'IC cleaned scalp EEG' over that from scalp EEG alone ( $p < 0.01$ , on means of 0.68 AUC and 0.64 AUC). Mean classification from a single selected ICA channel remained higher than IC cleaned scalp EEG data (0.70 AUC to 0.68 AUC), but this was not significant ( $p=0.15$ ).

From this, we conclude that selected ICA data or selected IC cleaned scalp EEG data gives better accuracy in classifying vision from EEG data using this kind of SVM protocol.

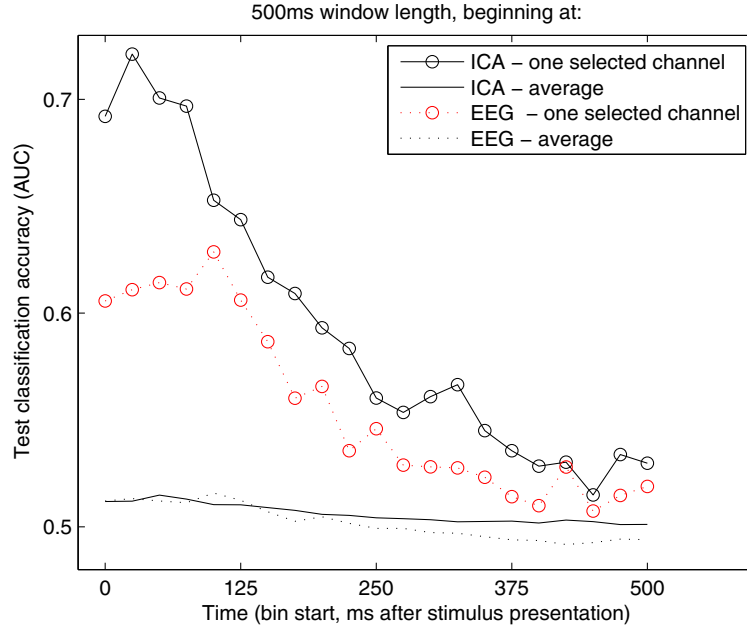


Figure 3.6: With use of a ‘sliding window’ of the time of object stimuli, we can see that trial labelling accuracy depends on being close to stimulus presentation. Using selected IC data (black circles) data from a 500ms window starting at 25ms (that is, data from 25–525ms post-stimulus appearing) gives 0.72 AUC, and data from 425–925ms will give only 0.52 AUC, suggesting less trial predictive information in the IC in that time bin.

### 3.1.7 More precise temporal profiling with a sliding analysis window

After demonstrating that some accuracy is possible at single-trial classification of ICA EEG data from visual object presentation when using all data within 500ms, we now attempt a training with many shorter trials beginning at different latencies compared to the object presentation. This allows us to examine both which scalp channels and ICs can predict visual state and at what times they have relevant information.

We continue training classifiers with 20 example trials (4 ‘positive’ trials in which object is present, 16 trials in which no stimuli is present) and classifying a further 5 unlabelled trials. The previous experiment used a feature vector of 128 datapoints (that is, 500ms of data from this channel at our 256Hz sample rate) beginning exactly at the onset of visual object stimulus or a non-time-locked trial in which no object was shown. We now attempt using only 50ms of data at a time. With labelling only 50ms of data from a trial as ‘positive’ and considering the first 50ms after object onset, then only the next 50ms, and so on, we can examine the performance of these classifiers when using data from each time point within a trial. We show sliding windows results in Fig 3.6 and Fig 3.7.

Even with training data consisting of 20 single-trials of just 50ms, we find above-chance classification and the accuracy is considerably greater using data from 75-175ms into the trial than at other timepoints.

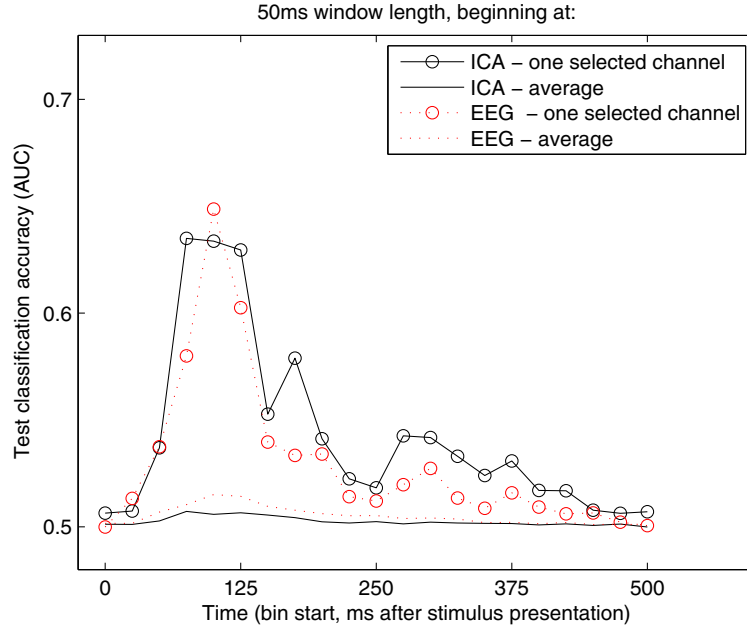


Figure 3.7: Reducing the size of this ‘sliding window’ so that only 50ms of data is used for training models rather than 500ms, we can see that data from 75–125ms is sufficient to give 0.64 AUC accuracy. Accuracy falls off sharply around this period. Data from 50–100ms are all at chance, suggesting no useful information for our classifiers at this time. Data from after 400–450ms are also back at chance.

If we examine the classification performance using each data source (each electrode and each IC) in more detail as we move this sliding window (Fig 3.8), there is typically a single IC with higher classification accuracy and the peak accuracy is around 50–175ms.

With these sliding windows, it could be that low-level visual properties like luminance contribute to the changing accuracy over time. Profiling the predictive data source could inform this.

### 3.1.8 Properties of the ICs

ERPs of a high and a low predictive power component are compared in Fig 3.9. The IC activations were mapped to spatial locations on the head (Delorme and Makeig, 2004). This mapping is shown on a standard 3D head model, for both a high and a low predictive IC. Also shown is the average ERP and the trial-by-trial ‘ERP image’ heat-map. The ERP image is a coloured plot in which every horizontal line shows activity from a single trial (Makeig et al., 2004). In this way, the variation in individual trials can be shown alongside the averaged ERP.

Here, the ‘high predictive’ IC has a complex activation pattern over time, and has a prominent decrease in IC activation from 50-100 ms after object presentation. In contrast, the activation of the ‘low predictive’ IC is not consistent across trials and does not have consistent timing in relation to the stimulus onset. The large ( $9 \mu V$ )

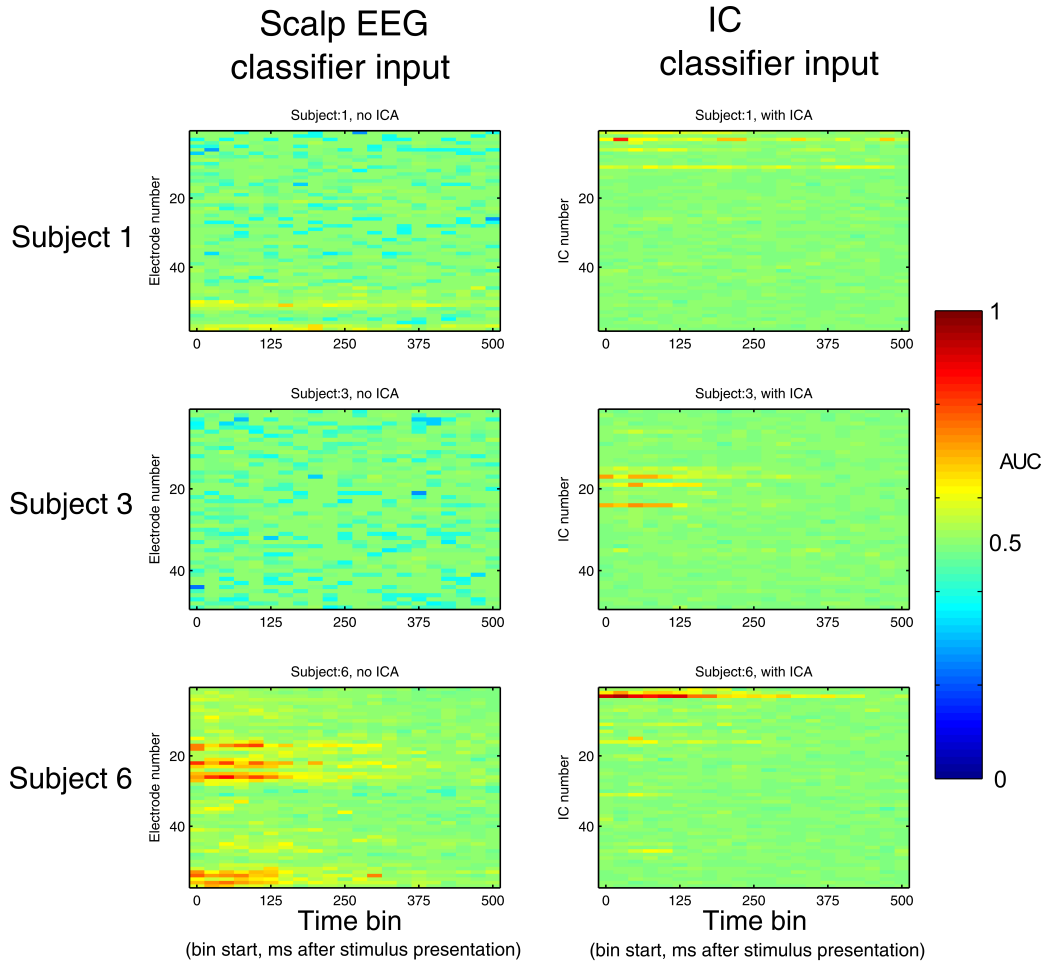


Figure 3.8: A heatmap plot of the classifier performance (AUC) for 3 participants using every electrode (left) and IC (right) as classifier input data, when using data from different timepoints in the trial. Outside the first 150ms, accuracy is low. Peak accuracy is higher in IC data, with our classifier better able to separate classes in that data type.

activation, alternating behaviour over time as well as IC localisation across the eyes, is consistent with this IC capturing signal from the EEG that corresponds to horizontal eye movements muscles.

To assess the range of IC spatial localisations, we ranked all ICs by object predictive power and plotted the spatial representation on 2D head-maps (Fig 3.2). There was a clear pattern of the five best ICs – and all ICs with AUC above 0.62 AUC – were found to be located in the occipital regions. The components with the lowest predictive power often mapped to spatial location of a single electrode, with these ICs likely capturing the activity of artefactual electrode noise (Onton et al., 2005).

### 3.1.9 Classification performance on each image

We report the success rate of classifying each object in Fig 3.10. This shows the classification performance on each of the 50 objects used as visual stimuli using the selected EEG and IC data, averaged across the 7 participants. Highest accuracy was using the selected IC data on the 'log.jpg' with 0.89 AUC.

### 3.1.10 Single Classification of 'Any Object Present' vs 'All Objects Absent'

In investigating the use of SVM classifying EEG data, an essential concern is evaluating what data can and should be used for classification. In order to inform this, we performed some initial classification assessments to give an indication of what might be possible.

Given our data was recorded in a visual stimuli presentation task, an initial simple target is finding the accuracy at which we can distinguish trials with any stimuli presented from trials with 'blank' stimuli, from intertrial time where the participant viewed a blank screen.

The data source, processing and classification procedure for this was as described in the Methods chapter. This was amended to examine all objects by taking 'positive' labeled data to be any of the 50 objects trials, with 'negative' data from a randomly chosen equivalent chunk of intertrial time where the screen had been left blank. Each training and testing example used 500ms of data from one electrode at a time, beginning at stimuli presentation time for positive examples. Training was performed using 40 example trials and tested on 10 further examples. Five-fold cross validation was used to improve consistency, and results were averaged. This was repeated for all 7 subjects, with presented results showing the average and variation across subjects.

From this, we can see highest data classification accuracy of 0.88 AUC when using data from a selected IC channel for training and testing data (shown on the right of Figure 3.11). Looking at most data sources, 500ms of data from a single channel of neither EEG nor IC data reveals much to distinguish visual stimuli vs blank trials in this protocol, as grand average accuracy is close to 0.5 AUC (leftmost bars on Figure

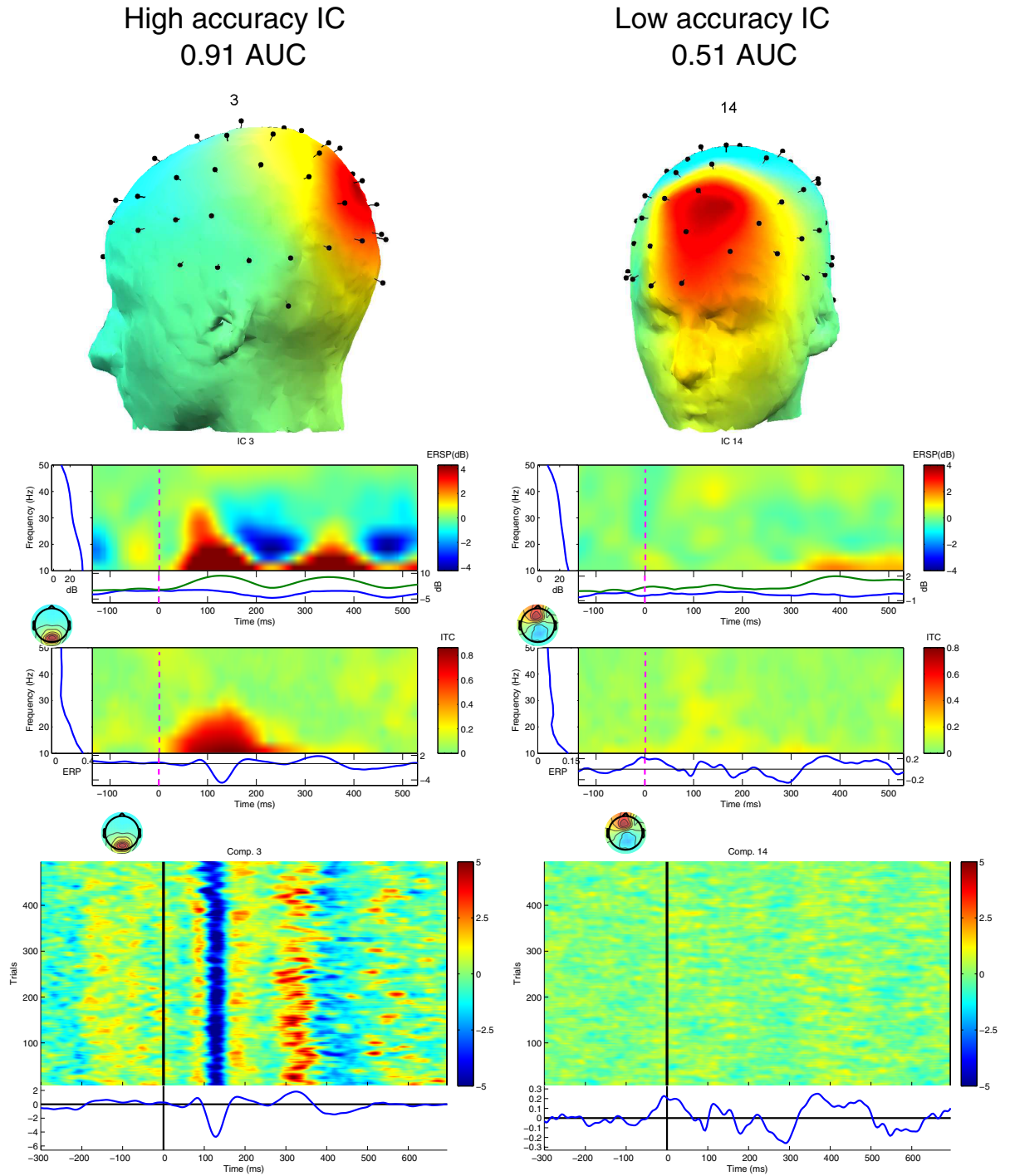


Figure 3.9: Individual IC activation details. Activations of two example ICs in participant 6 – component 3 (left) and component 14 (right). Top: standard 3D head models with EEG electrode positions as black pins and colours showing the spatial distribution of two ICs, where red areas of the head indicate higher EEG  $\mu V$  in this area when this IC is active. In this example, IC 3 is an occipital component and IC 14 has an activation profile consistent with eye movement. Bottom: corresponding IC ERP-image plots showing the 250 individual object presentation trials, with red indicating higher activation of this IC at that time. The lower plot is the standard IC ERP averaged over 250 trials.

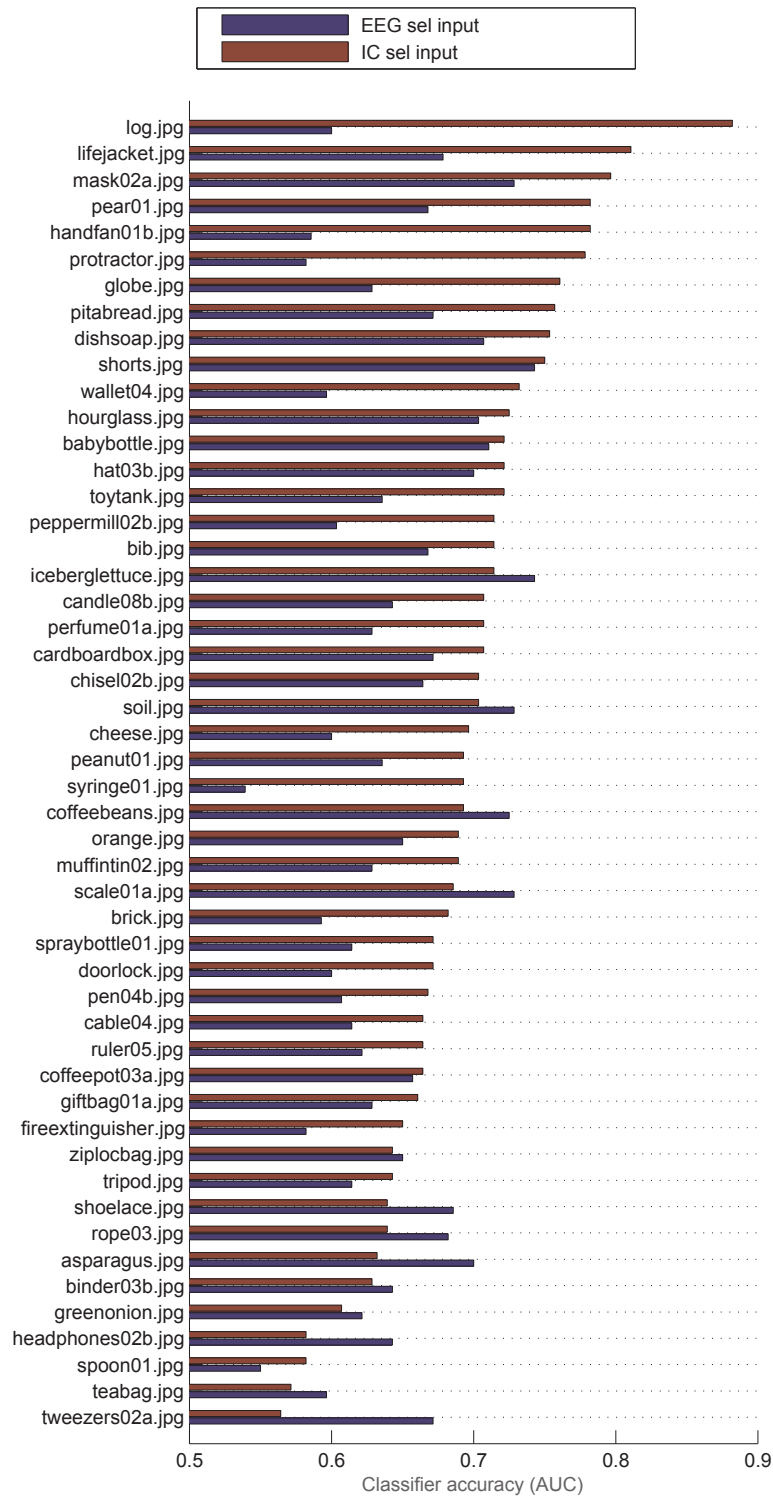


Figure 3.10: Classification performance shown by each object of our 50 visual object stimuli, sorted by IC AUC. This is averaged across the 7 participants, using the selected IC channel data (shown in dark red) and selected EEG channel data (dark blue). Single-trials with pictures of the log, life-jacket and mask were labelled correctly at high accuracy, and the IC outperformed scalp electrode data for most objects.

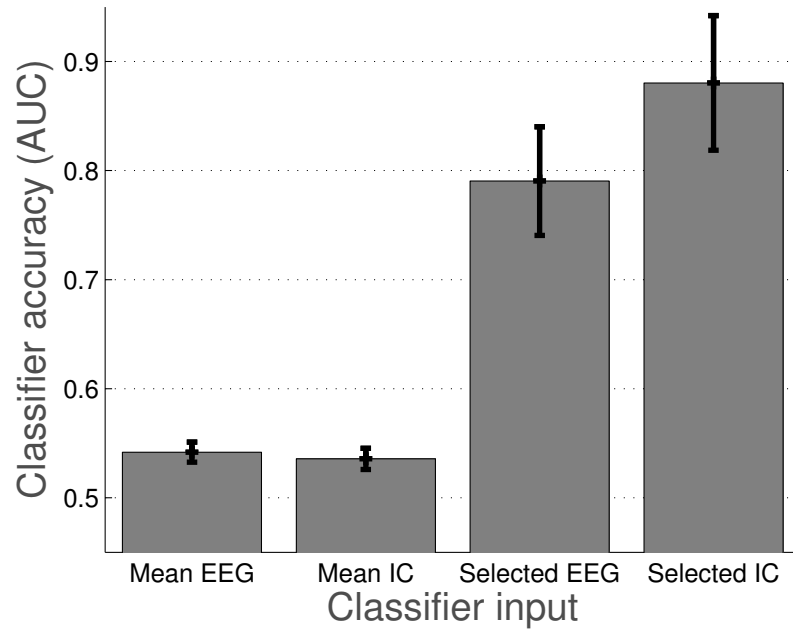


Figure 3.11: Classification accuracy of various data sources when training and testing on trials from all objects labelled as 'positive' and blank intertrial time labelled as 'negative'.

3.11). However, there are a few channels that appear to have greatly more information towards this, and using the channel selection procedure, average accuracy is raised to 0.79 AUC for EEG and 0.88 AUC for selected IC channels.

Classifying all objects vs absence of any objects informs us that most channels do not contain useful information in this protocol, and also gives us bounds of the high accuracy of a single selected IC in this easier classification task. This was expected to be easier than single object classification as greatly more training data exists (with data from all 250 objects presentations, 5 x 50 objects, rather than the 5 trials for examining single objects). As might be expected, accuracy is higher with the additional training examples, but object-specificity is lost.

### 3.1.11 Object category classification

#### Examining possible object categories

We continue to investigate feature design in classifying visual state from EEG data. In order to examine what can be learned, the choice of which examples to label as 'positive' and which as 'negative' is vital. When this is 'all objects' vs 'blank trials', as in 3.1.10, we can see the peak accuracy is from using data from a selected IC channel in that protocol.

One alternative labelling would be to examine results of a single classifier for each object – 'each object' vs 'blank trials'. From this, there is an advantage in that the salience of each object might be noticed. That is, if the neural response (as detected



<b>Natural food image category 'positive' class label</b>	
Name	Object ID
'greenonion.jpg'	7
'asparagus.jpg'	11
'pear01.jpg'	36
'iceberglettuce.jpg'	38
'peanut01.jpg'	41
'orange.jpg'	46
<b>Artificial household / tools category 'negative' class label</b>	
Name	Object ID
'ruler05.jpg'	21
'chisel02b.jpg'	22
'protractor.jpg'	26
'pen04b.jpg'	27
'spoon01.jpg'	32
'fireextinguisher.jpg'	42

Table 3.1: Object categories - a subset of 6 objects chosen for comparing possible categories. We show the filename and the Object ID, used for reference. These category labels agree with those used by the image database curators Brodeur et al. (2010).

with EEG) is substantially different between objects, classifiers may be trained to pick up these differences in the signal.

A different set of information could be picked up by training classifiers on different categories of objects – such as 'natural objects' vs 'artificial objects'. Previous studies suggest relevant categories to consider. Examining the neural response to different objects suggests natural categories of: animate bodies and faces, inanimate natural objects, and inanimate artificial objects (as discussed in Kriegeskorte et al. (2008)). Given the set of objects we used from the 'Bank of Standardized Stimuli' (Brodeur et al. (2010)), obvious categories appear to be natural food items (such as 'pear', 'orange', and 'peanut') and artificial household items ('ruler', 'chisel', 'spoon').

Can we use this machine learning on EEG approach to separate the neural response from different categories of objects? Is this difference larger than that of 'all object' vs 'no object' condition?

We investigate this by amending our classifier training and testing protocol to now label a set of 6 natural food images as the 'positive' examples and a set of 6 artificial household items as 'negative' examples. From the 'Bank of Standardized Stimuli' images used, we take the category label that the authors report in the 'Modal category' in the supporting information spreadsheet (Brodeur et al. (2010) BOSSNORMS.xlsx, as found on <https://sites.google.com/site/bosstimuli/download>). The food items have the label 'food', and the artificial household items have the label 'Handlabourtools' and 'Kitchenutencils'.

## Object category classification results

First, we train classifiers as before, with object stimulation trials and blank stimulation trials, but with all objects from within a category taken as 'positive' data. We find that for both the natural food category and the artificial objects category, selected IC data was able to correctly classify trial identity at higher accuracy than single objects (means of 0.936 and 0.938 AUC respectively, as shown in 3.12).

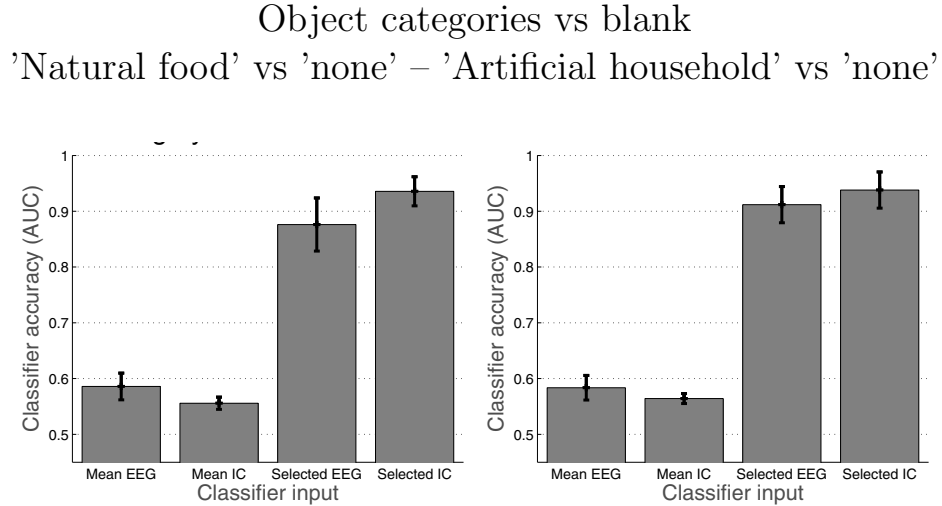


Figure 3.12: Classification of object categories vs no object trials.

**Left** - 'Natural food' category. **Right** - 'Artificial household' items category.

The increase in classification accuracy here compared to single-object classification model could reflect the additional training examples, and suggests that the additional training trials did indeed have constructive information toward accurate trial discrimination.

Additionally, we also examined the ability to determine a trial from one category from that of another. This gave greatly reduced accuracy (down to 0.72 AUC for the best selected IC source, see Fig 3.13).

Intriguingly, the models trained to distinguish from different object categories performed at much lower accuracy than those distinguishing object categories from blank trials. This suggests that, in the context of this protocol, the brain response to a food item object presentation is more similar to that of a household item object presentation than that of a blank trial.

Initial investigation of training models to classify trials from 'duelling' single objects vs other single objects gave low accuracy (0.55 – 0.65 AUC) and low consistency. This could be due to the relatively few training examples of only 5 trials per object for both positive and negative examples.

## 'Natural food' vs 'Artificial household' object categories

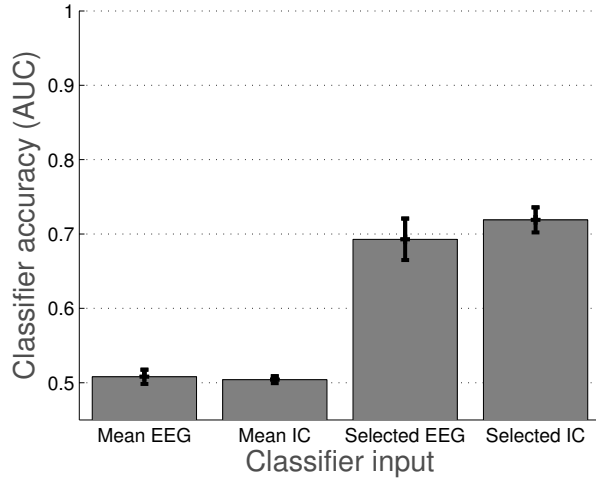


Figure 3.13: Classification accuracy of models trained to separate object trial categories of 'Natural food' category vs 'Artificial household' object categories.

### 3.1.12 Classification across subjects

So far, classification training and testing has been performed on a subjectwise basis, where only data from that subject is used. Results from using single selected IC data for training and testing indicate that an occipital alpha dominated IC gives high accuracy in several subjects.

Rather than performing this source selection procedure on each subject, can we identify the highest accuracy IC across all subjects?

Comparing ICs across subjects is not trivial, as the ICA algorithm will likely find quite different source components in each subject. This is partly due to the varying neural patterns, resting EEG activity, head shape and skull thickness that naturally varies across people. Nonetheless, sensible comparisons of ICs can be made (Beckmann et al., 2009), even using ICA across all subject data (Hyvärinen, 2011). There are properties of identified ICs that lend themselves to concise description, such as spatial distribution and spectral properties, which may be used to place ICs in groups that can be identified across subjects.

One approach is to examine the properties of these identified ICs in all subjects and group the ICs together in clusters, putting ICs where the similarity of these properties is highest in a group together. This can be done using k-means clustering on group properties of the EEGLAB STUDY data structure (Onton et al., 2006).

With around 70 ICs identified in each of the 7 subjects, there are around 490 ICs total. Running k-means on this, we can group ICs by spatial similarity.

### 3.1.13 Cluster properties

Using k-means to group ICs together into clusters, we find scalp maps, ERPs and ERP-images as shown in 3.14. Using this simple grouping method, some properties

## 2D scalp maps, showing average IC cluster location

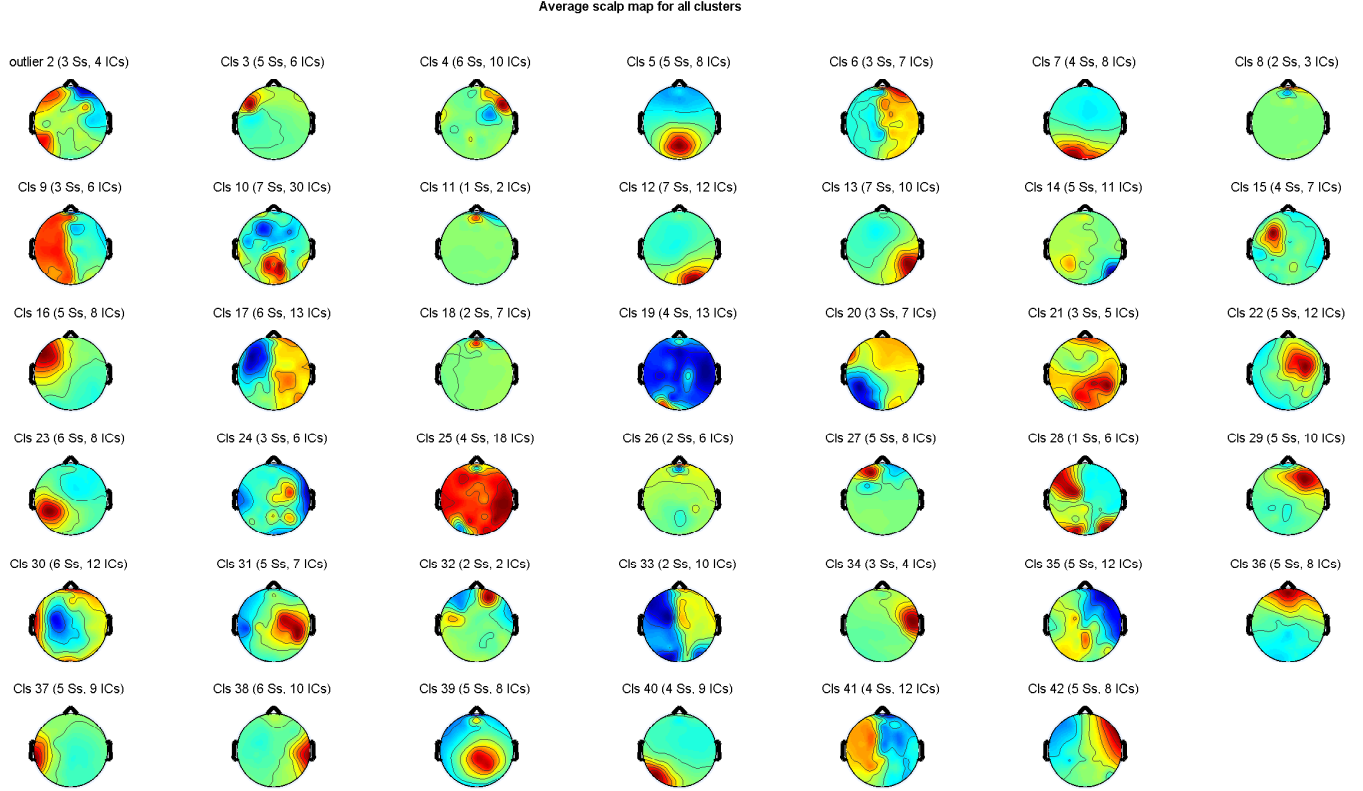


Figure 3.14: 2D scalp maps, showing average IC cluster location. Cluster groups were found using k-means. For example, 'Cls 5' here is cluster number 5, shows 8 ICs from 5 different subjects, and represents occipital ICs. 'Cls 1' is taken to be the set of all ICs.

are placed together. Cluster 5 has 8 ICs, all roughly smooth Gaussian in shape, and localised centrally and occipitally. Some clusters are more erratic in appearance (cluster 10), but many show a consistent locus.

These cluster plots show that the k-means grouping does indeed appear to correctly cluster ICs into groups of somewhat similar spatial distributions. We also show the ERP activity of those IC groups in 3.16.

Can this grouping be used to identify high-predictive ICs? This was investigated by performing classification as before, classifying either 'natural food' images vs no image, or 'natural food' images vs 'man-made object' images. Many clusters show fairly low accuracy. This is in agreement with them being a group of non-relevant ICs, so giving low trial predictive accuracy. Cluster 35 gives the highest accuracy, at 1 AUC on separating 'natural images' vs 'no image' trials. The scalp maps, as shown in Fig 3.14, demonstrate that this cluster is distributed tightly around the eyes, and has activity consistent with that of averaged eye movement.

Intriguingly, when the classification task is changed to distinguish data from one

## IC members within two clusters

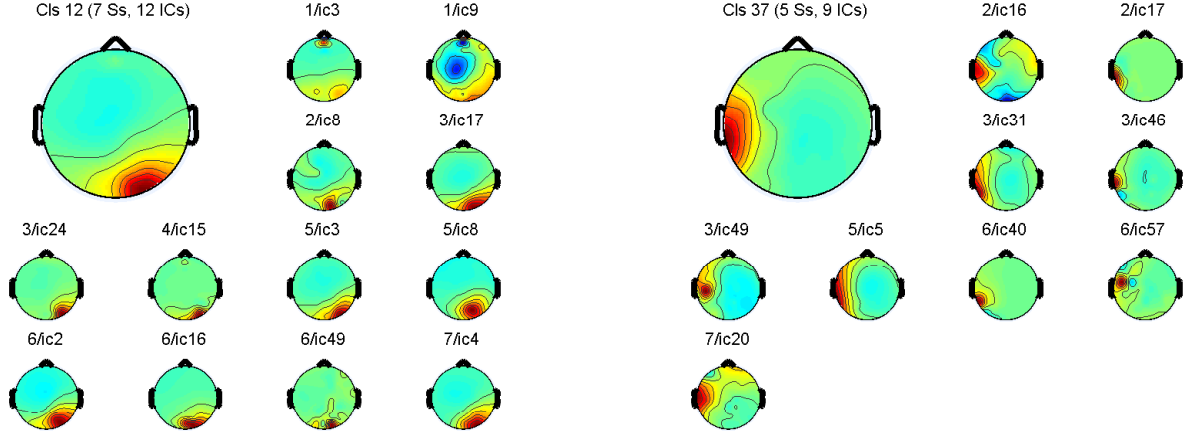


Figure 3.15: 2D scalp maps, showing each member IC of two example clusters. The spatial distribution is relatively consistent. **Left** - cluster 12 members are mostly right-occipital. **Right** - cluster 37 member ICs are mostly left-temporal.

set of objects to another set of objects, the accuracy from this eye-movement cluster 35 drops greatly, but the accuracy of an alpha-occipital central cluster (cluster 5) remains relatively high at 0.733 AUC.

This comparison across subjects allows us to group together similar ICs and so make statements that hold across subjects. Given the high-predictive nature of some clusters in many subjects, this cluster identification could be used to replace the selection process that was used to find high-predictive ICs in earlier experiments.

## Average event-related potential (ERP) of each cluster

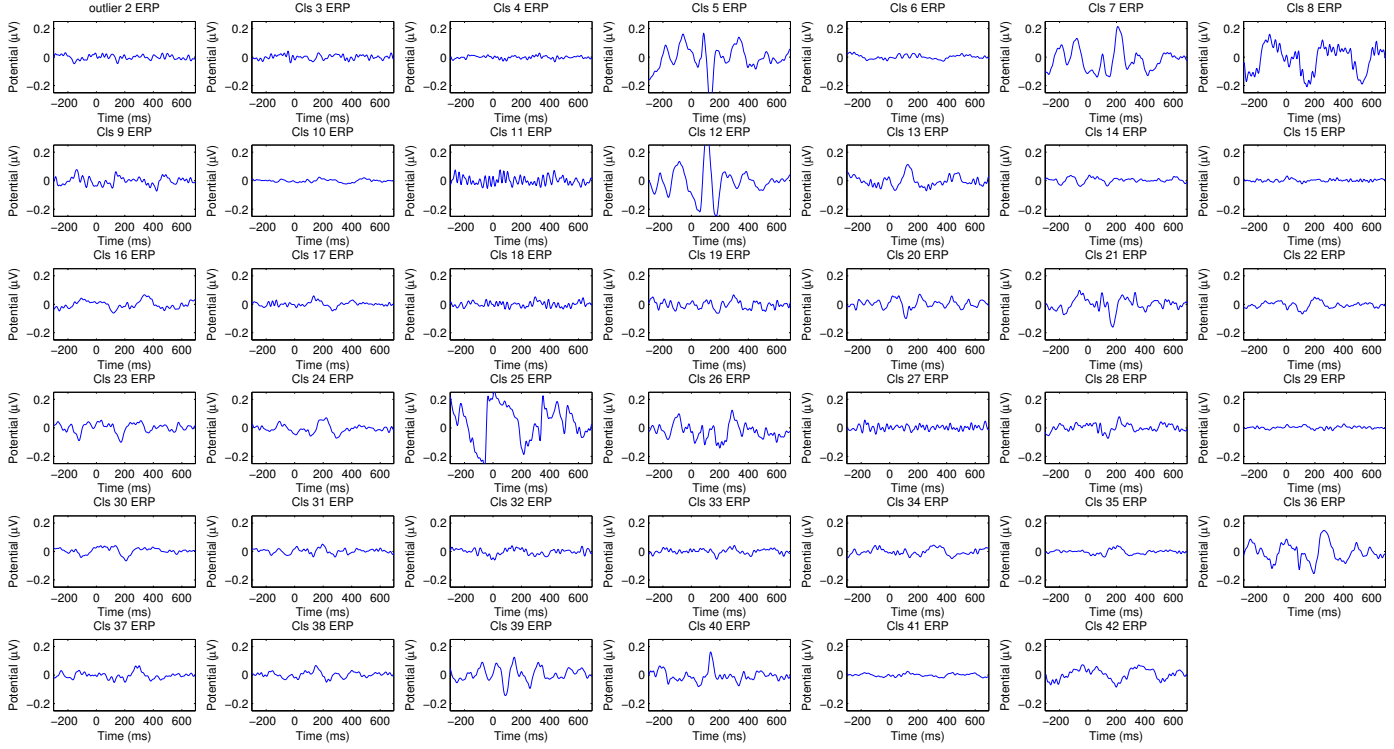


Figure 3.16: Average ERPs from each cluster. Clusters are those shown above in 3.14. While these are recomputed to uV, this is after data has been normalised.

### 3.2 Visual classification discussion

We proposed a method for classifying visual state from EEG using machine learning. We demonstrate that this can distinguish data from visual object presentation trials from data without object presentation at above 87% accuracy.

We found visual classification accuracy well above chance in 5 of 7 participants when first using input from all channels of EEG data. Mean accuracy and robustness was increased when using only a selected subset of 1 EEG channel of data for input, rather than all at once. Further, we found that training classifiers using input data from IC activations gave higher accuracy.

We can identify that the ability to label single-trials is dependant on information in the selected data channel at 75–150ms into the trial (Fig 3.7).

#### 3.2.1 Improving accuracy with selected input

The finding that input of a subset of a single selected data channel outperformed using all channels together deserves consideration. This may be due to classifier ‘overfitting’ – where the model parameters fit the properties of the training data too closely and so

### Classification accuracy using data from each cluster

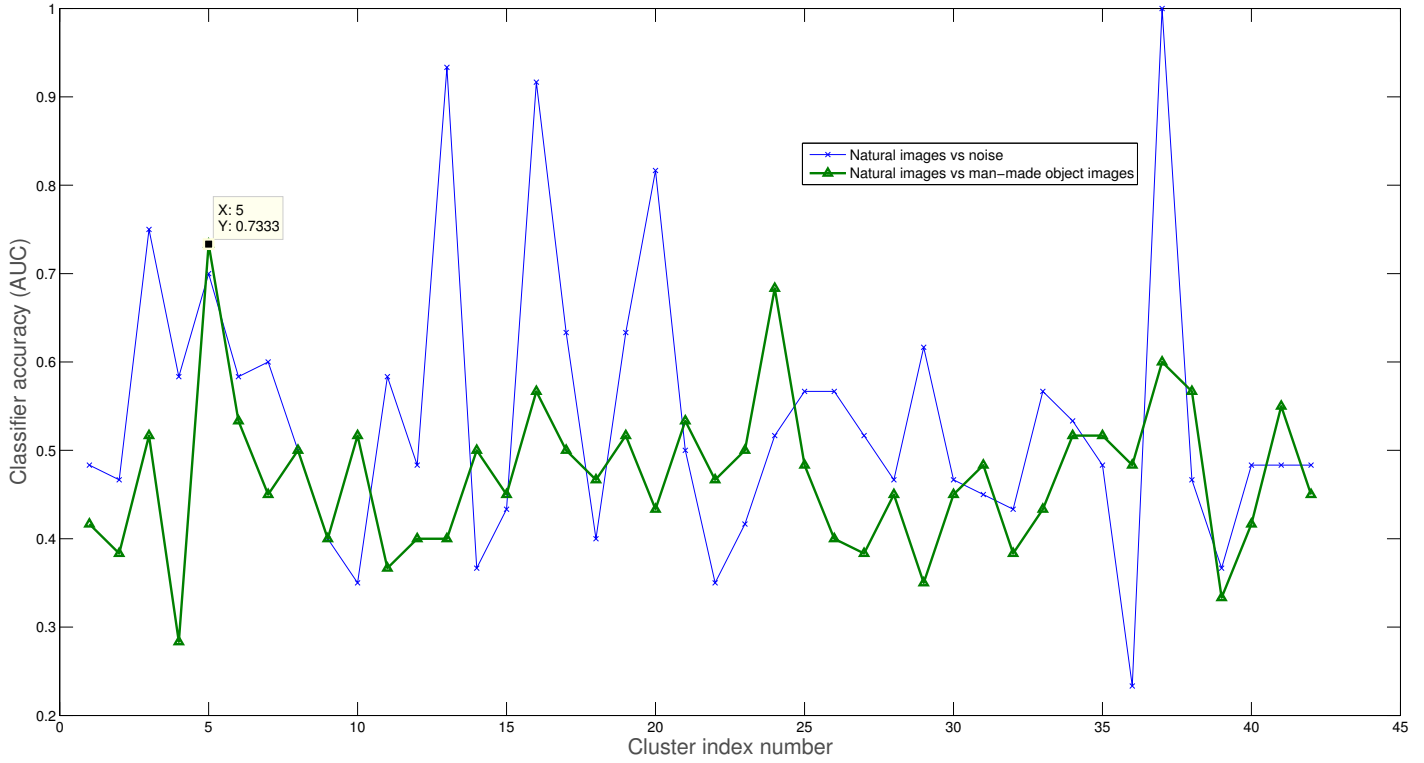


Figure 3.17: Classification accuracy using data from each cluster. Classification of 'natural food' images vs 'man-made object' images (dark green) is generally low, but cluster 5 shows higher accuracy of 0.733 AUC. Classification of 'natural food' images vs no image (blue) is notably higher in several clusters. Highest accuracy is when using data from cluster 36.

do not best generalise to classifying new data (Babiyak, 2004). The overfitting when using all data could be due to suboptimal model parameters of the training error cost parameter ' $C$ ' and the radial-basis function kernel parameter ' $\gamma$ '. However, these were the best values provided by a parameter search.

Generalisation was improved by using appropriate dimensions of the classifier input data. Using one channel of input data rather than all  $\sim 49$  channels gave somewhat increased mean accuracy. Notably, this considerably increased reliability (from 0.52 AUC to 0.65 AUC), with classification accuracy well above chance. This could be as a result of our implementation, but does indeed still show that a single selected channel of EEG input data can be a consistent and concise source of input data for high-accuracy classification.

#### 3.2.2 The use of ICA

When classifying presence of visual images, using classifier input of the IC activations from a single selected IC gave highest accuracy in all seven participants at 500 ms window. Although the average IC input classification was approximately equal to the

average single EEG channel input, those ICs that performed best on the initial 10 objects continued to outperform EEG channel input in further test data (0.70 AUC using selected IC channel to 0.65 AUC in the selected EEG channel). This corresponds to a significant increase in timepoints correctly labelled.

In trying to improve performance of neuroimaging classifiers, there are two domains commonly focussed upon: feature extraction and the classification itself (Farquhar and Hill, 2006). EEG data is intrinsically noisy, contains highly correlated features and has much variance both between different participants and within the same participant over time. This suggests that considering decomposition of possible EEG sources could be useful when assessing features to extract from EEG (van Gerven et al., 2009). ICA does this by decorrelating the inputs and attempting to minimise mutual information in forming the components.

In previous studies, ICA has been shown to improve performance of a classifier in a simple auditory task (Hill et al., 2004). To our knowledge, the present report is the first to document ICA giving such an improvement in a visual object classification task.

ICA appeared to separate sources of noise into some ICs, and sources of task-related neural activity into other ICs. This resulted in the single-trial classification performance in many ICs being mediocre, but a few ICs giving higher accuracy than that of any EEG channel. Thus, ICA may be thought of as concentrating source signals into de-noised IC channels (see Fig 3.5), and so our SVM classifiers can give higher accuracy using this input. Examining IC input also has the advantage of assessing a source-space deconvolution (Bell and Sejnowski, 1995).

This view was also supported by an improvement in trial classification accuracy from EEG data where probable-artefacts have been removed (see Fig 3.5).

Within participants, many EEG channels had relatively similar predictive power. While the magnitude of the averaged ERP might be quite different on electrodes across the scalp, at a single-trial level much of the signal present at any electrode is also present on neighbouring electrodes (Luck, 2005) albeit at a different scale. As the SVM classifier here is separating classes in a high-dimensional hyperplane, it can remain sensitive to small (but consistent) separating features. Thus, features that would not be obviously visually apparent on a plotted ERP can still be captured.

We suggest that the task predictive power of an IC has the potential to inform theories of visual cognition. For example, an IC that is localised in neck muscles, and also gives low accuracy when used to classify the task, would be an unlikely target for further investigation. In contrast, an IC that is spatially localised in visual cortex, and that gives high accuracy of classifying the visual state under specific conditions, may be a fruitful target to profile.

### 3.2.3 Visual object processing

Although EEG studies have shown stimuli-specific ERP separation within 150 ms (Thorpe et al., 1996; Johnson and Olshausen, 2005; Mouchetant-Rostaing and Giard,



2000), there appears to be some uncertainty as to whether these differences represent low-level visual properties or higher-level cognitive categorisation of visual stimuli. Related work in macaque monkeys using multi-unit neuron cell recordings seems to indicate clear object- and category-specific information in primate inferior temporal cortex at 100-125 ms after presentation (Hung et al., 2005). Cell firing was used as input for a classifier to correctly identify presented object stimulus at 70% accuracy using only the spiking behaviour of a few neurons in inferior temporal cortex.

This neuron spiking data strongly indicates that the brain activity is present at these time-scales, but that, of course, does not necessarily mean that we could also observe this within EEG data, which would allow more probing in humans. Clearly, the ability of averaged ERP EEG analysis alone is limited in addressing this (Makeig et al., 2004; Rousselet and Pernet, 2011). We suggest that our approach of ICA and SVM machine learning is a more suitable one.

### 3.2.4 Cognitive implications and future work

The identification of EEG sources that have robust task-related predictive power would allow cognitive science experimental designs that target that specific IC. Getting a profile of IC localisation, task dependence, temporal and spectral properties may be much more accessible in probing possible underlying neural processes in greater detail than using EEG data alone. Observing increased accuracy when using IC activation input suggests that they are a promising target, although the extent to which specific ICs might relate to underlying neural processes is not yet well known (Onton et al., 2006).

As expected, artefact related ICs had low predictive power to discriminate stimuli. Some occipital components did vary with stimuli, but the best IC was one found in one participant that was localised in the left temporal area. Further testing could investigate whether this was related to the naming of the presented object.

Several of the components seemed to be related to eye movements, muscle artefacts or electrical noise, as identified by back-projected location and activation properties (Delorme et al., 2007). Nonetheless, we have shown here that several ICA components have a greater degree of information content that can be used to predict perceptual processes than unprocessed EEG data had. Here we have shown that a selected IC can classify better than EEG when the classifier is given 500 ms of each. It might also be informative to profile the task-classification performance using data of different time-points. For example, is there one IC that better captures the N170 'ERP component', typically associated with faces and certain other objects being presented (Joyce and Rossion, 2005), and another IC that better represents the oddball P300 ERP component (Polich and Kok, 1995)?

Some studies have presented evidence towards the conclusion that there is different information in the EEG around N170 and the P300, and that what can be predicted from each stage is indicative of underlying neural processes. The component often

referred to as N170 has been shown to have information that can be used to discriminate faces from cars (Philiastides and Sajda, 2006). This demonstrates that single-trial EEG can be used to distinguish between visual percepts. In Philiastides and Sajda (2006), visual choice decision-making between cars and faces was examined. Here, we used a larger selection of different visual objects and, using data from a few single-trial training examples, attempt to use SVMs to label which object was present in that trial.

Other studies of visual perception have considered the similarity of fMRI activity in inferior temporal cortex in response to objects of different categories (Kriegeskorte et al., 2008). The current study used 50 pictures from the BOSS image database, all of which are considered to be in the ‘food’, ‘hand-labour-tools’ or ‘kitchen-utensils’ categories.

Fast, high accuracy classification of visual state also allows real time detection within a recording session, and so experiments involving EEG feedback or manipulation could be performed. Do any IC have activations associated with borderline visual percepts, or is the task-related representation we see here downstream of an all-or-nothing percept trigger?

These components – along with our SVM prediction models – can classify new trials of visual stimulation well, but could likely be used to probe other steps of visual processing and perception. We suggest that profiling the activity of these components identified by ICA in different tasks, quantified with machine learning, might be fruitful conceptual targets for future visual processing experiments. Might any of their classifiers respond to imagined objects rather than presented objects? Given several objects presented simultaneously to a participant, might some classifiers respond to the attention of the participant and some respond only to the immediate visual input, regardless of context? With a visual task of noisy stimuli where object perception is only sometimes reported, will some classifiers predict perceptual onset even before perception?

### 3.2.5 Limitations

The metric used for ‘prediction power’ gives an indication of task-related information, but it will also have some dependencies on the properties of the classifier used (Meyer et al., 2003). Nonetheless, the prediction power of those ICs that likely capture artefactual signal is low, suggesting that task prediction power can be a valid proxy for task-related information content.

Models trained on all channels of EEG data were not as robust as classifiers using single channels and failed to classify on two of seven participants. This suggests this classifier model was not fitting the signal-to-noise of this data well, and overfitting (Chang and Lin, 2011). Using one selected ‘best’ channel of IC data resolved this, but at the expense of not utilising all available input data.

It may be the case that ICA was particularly useful here, as little other preprocessing was used. Prior research suggests that use of tailored data extraction can increase

accuracy on classifying EEG data (Blankertz et al., 2006), where frequencies of interest, channels of interest and input ranges are manually specified. We suggest that in cases where the anticipated activity is relatively unknown, use of the more automated ICA might avoid misuse of tailored data extraction when it is not appropriate.

We considered classifying the presence or absence of single images. It could be possible to extend this to instead classify each image against the other images 'apple vs banana' rather than 'apple versus apple', or through use of a multi-class classifier, but we have not addressed this here.

### **3.2.6 Visual Classification summary**

We presented a method for labelling single-trials of EEG using SVM classifiers. Visual object presentation state could be classified at 0.70 AUC (more than 87% accuracy). Robustness was increased by using data from a single selected EEG channel. Further improvements in accuracy were found using data from a single selected IC channel.

We suggest that this method of machine learning and independent 'source separation' might allow detailed probing of information content within EEG data.

### 3.3 Results of Online Analysis experiments

#### 3.3.1 Stage I – training 'one-vs-many' object models

New data processing and handling Matlab code was produced so that LSL could import live EEG data in real time, store this data efficiently, train classifiers on this, and export this data in a way that is compatible with the EEGLAB data structure. With the LSL Biosemi interface running, the data continuously streamed to a Matlab object data structure. This was written to a 'chunk' in each trial, with the timing synchronised with Psychtoolbox screen blanking timestamps. With *make\_EEGLAB\_struct\_from\_chunks.m*, we convert this data into EEGLAB format to allow use of standardised EEG filtering, referencing and ICA tools.

For Stage I, the 'Visual Classification I' protocol from offline analysis was largely repeated. The participants were instructed to pay attention to the screen while everyday visual objects were presented (Stewart et al., 2014).

Changes to this protocol included shorter trial time (so that data from more trials could be recorded within a session), and the use of LSL online EEG data handling, rather than the use of Biosemi software to passively record EEG data to disk.

Rather than examining object trials vs no-object trials, we also switched to classifying target object trials vs data from trials from all other objects. That is, rather than classifying trials in which an apple is shown to trials where nothing is shown, we now classify trials with apples against trials with bananas, carrots, and blank trials.

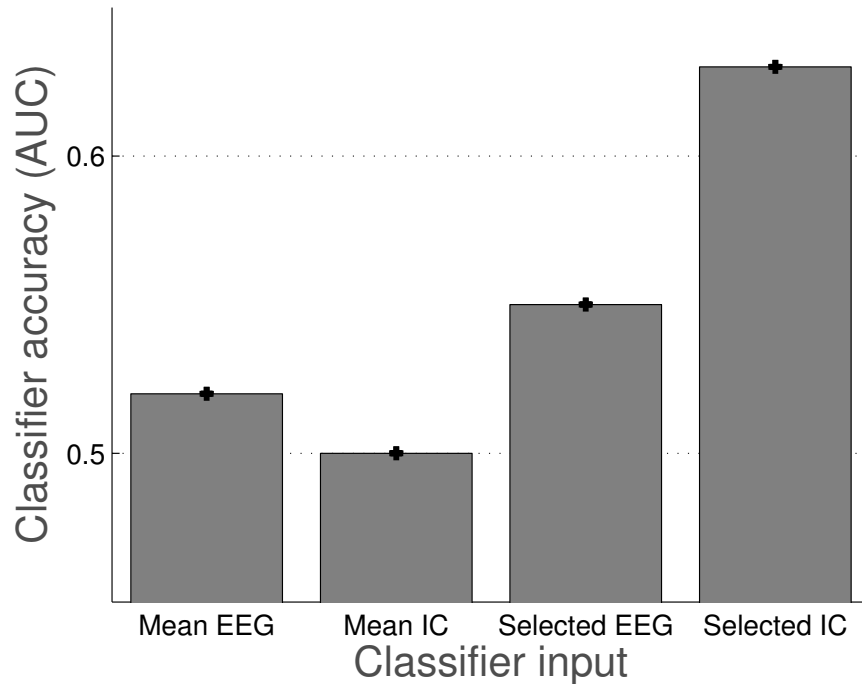


Figure 3.18: Classification results for 4 participants in online Stage I. Average accuracy was found to be when using data from the selected IC

With this analysis, mean accuracy was found to be lower here, using short object-object trials, than when using longer object present-absent trials in visclassica before. One-versus-many is of course a more difficult classification task, but we still found participants with mean 0.61 AUC accuracy at correctly labelling trials as 'apple' or 'object that isn't apple'. Again, peak accuracy was found using the selected IC data (Stewart et al., 2014).

<b>S1 AUC target object</b>		
	Best EEG	Best IC
Subject 1	0.57	0.61
Subject 2	0.57	0.57
Subject 3	0.54	0.57
Subject 4	0.58	0.56
Subject 5	0.55	0.83
Subject 6	0.52	0.54
Subject 7	0.55	0.67
Subject 8	0.53	0.70
	Best EEG	Best IC
mean	0.55	0.63

Table 3.2: Table: Mean accuracy of online Stage I object-object classification (AUC) on each participant, using selected classifier input data

The properties of the high predictive ICs could be examined in the experiment, and activity observed in near real time.

### 3.3.2 Stage II – classifying visual with and without attention

Previously, with single object stimuli, trial classification accuracy could be a result of trial-specific neural activity relating to low-level visual processes, neural activity from attending or perceiving this object, some intermediate, or some combination.

In Stage II, we move from examining single object presentation to two overlapping images. Now, with data from the subjects focussing attention on the target image (in the presence of a distractor image) and the same image but without attention on the target, we can train classifiers to predict the presence or absence of attention on each image.

Trial structure and example stimuli are shown above, in Fig. 2.9 on page 46. As a reminder of the setup, we show 150 trials of these overlapping images, with the 8 objects and a blank image shown in random order and combination, with each object being used as attention target or distractor. Attention target is prompted with a text screen between trials. Trials last 2-3s. At the end of each trial, the EEG data from that trial is saved from the LSL buffer to a Matlab array, along with specification of the two objects that were shown, which was the attentional target, and the subject's confirmation of attention focus.

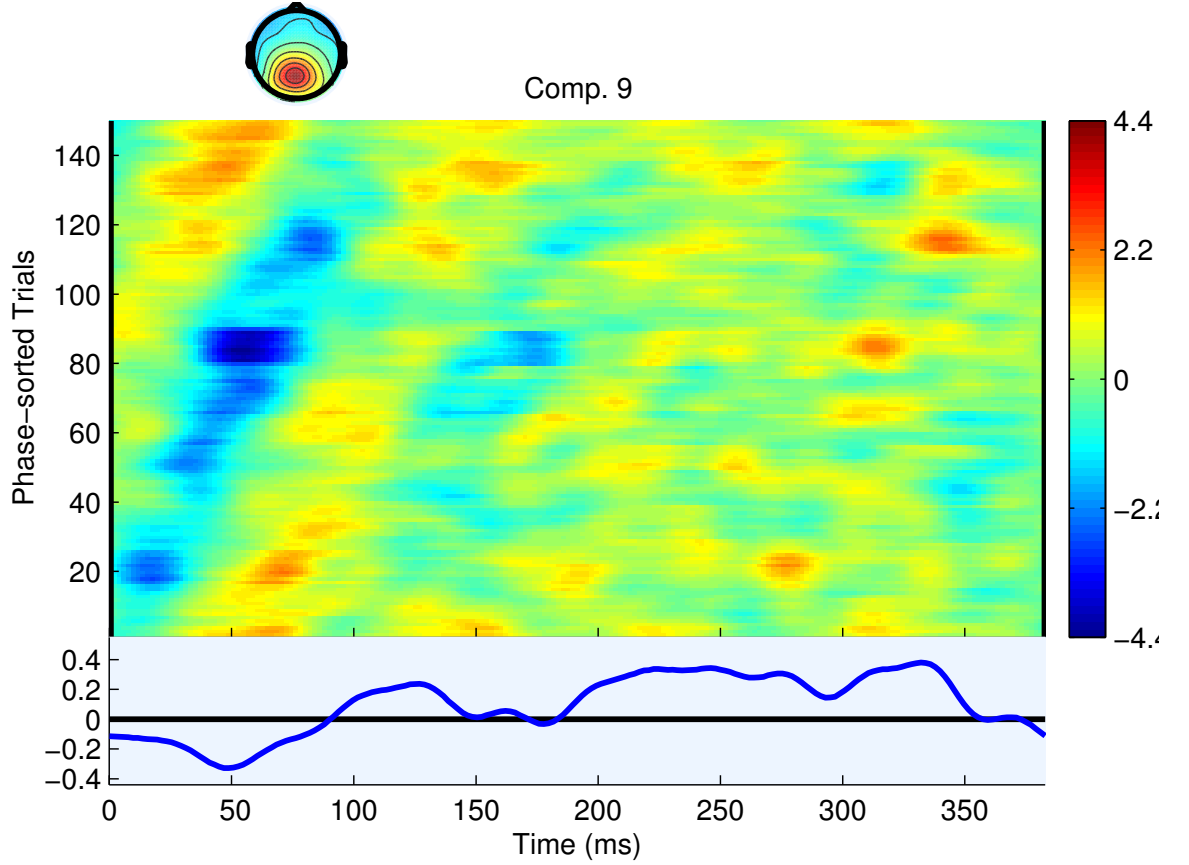


Figure 3.19: Activation details for the highest predictive IC in participant 10, in all 150 S2 trials. Location appears to be around the central occipital-parietal region.

### 3.3.3 Classification accuracy on ambiguous images, with and without attention

First, we used models trained on stage I serial object presentations, and assessed the accuracy of predicting trial label for stage II.

With the training data from stage I, we have 150 single-object presentation trials, with around 15 presentations each of 8 different objects, along with 'dummy' blank images. For each object, there is then around 15 positive examples per subject, and a maximum of 135 negative examples. We trained models using 400ms of data beginning at time of image presentation. At 512 Hz sample rate and 400ms analysis windows, each example is a vector of around 200 datapoints. We have 32 channels of data for both EEG and the ICA transform, so a data matrix of  $2 \times 32 \times 200$  datapoints for each of the 150 examples. The ICA mixing matrix was computed once, and this was done at the end of stage I.

For stage I, 'positive' data was target object being present, and 'negative' data was data from any other trial. For stage II, we use these already-trained models, but investigate the performance of these classifiers on new test data, where we can test

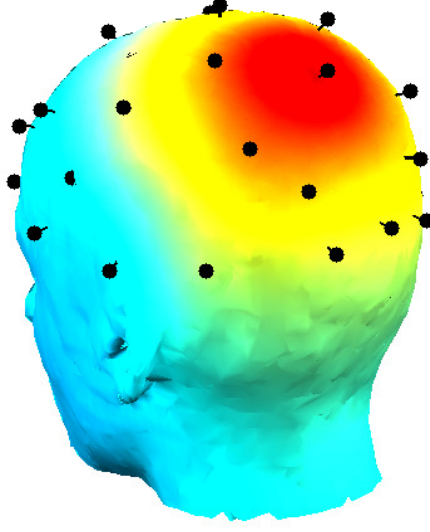


Figure 3.20: A 3D headplot of IC9 from participant 10, showing the distribution of this IC over the scalp.

data from different trial states from stage II, such as comparing attention on object, or attention not on this object.

We show the results of this in Table 3.3. Intriguingly, even with the models being trained on prior single-image presentations, we have 0.73 AUC when the target object is shown and is the focus of attention. When the attention is moved to another image, accuracy drops to 0.58 AUC. We suggest that this difference -  $\Delta$  AUC - on attentional focus indicates that some of the classification accuracy may be related to attention-dependant neural processes, as the same image (pixel for pixel) is shown in each case.

<b>Stage II Classification accuracy with models trained on s1 IC data</b>		
Stimuli displayed	Attention on	Prediction accuracy (AUC)
Target object shown with distractor	attention to target	0.73
Target object shown with distractor	attention to distractor	0.58
Distractors only	attention to distractor	0.61

Table 3.3: Table: Mean accuracy of online Stage II object-object classification (AUC) from each subject, using selected classifier input data

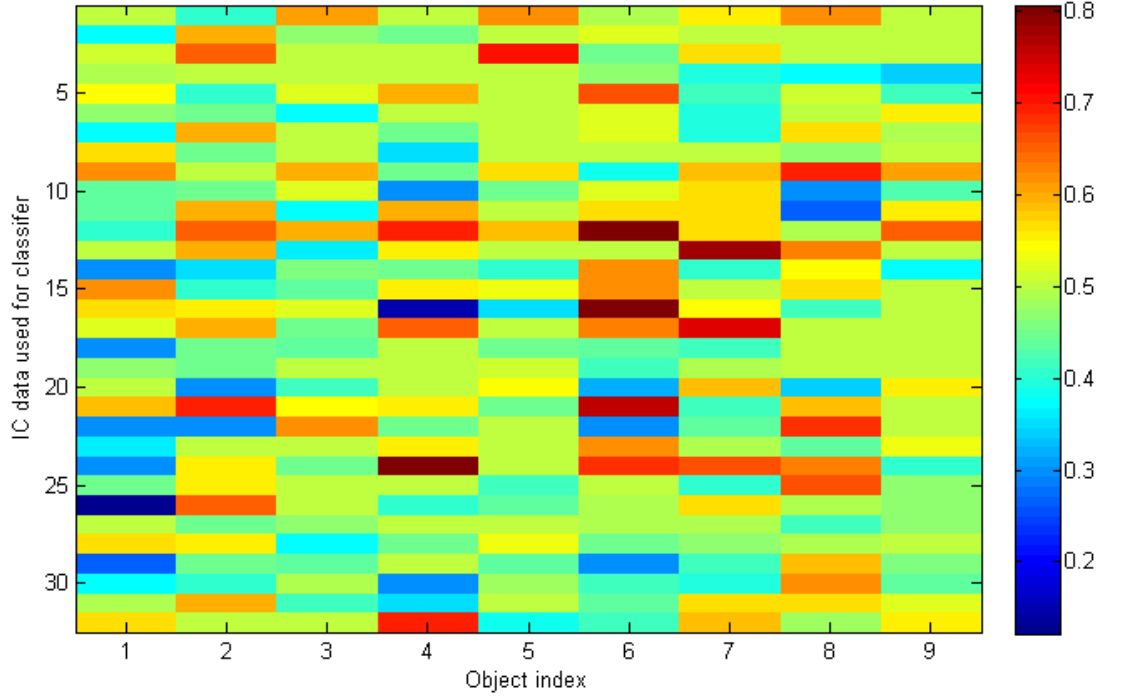


Figure 3.21: The accuracy on 9 different objects. Here, IC 12 is shown to have average 0.61 AUC over these 9 objects, with highest accuracy on object 6, protractor.jpg

### 3.3.4 Attention-specific classifier properties

With this stage II data, we can then attempt to train attention-specific classifiers. That is, a classifier that is not selective to the displayed stimuli, but that is sensitive to attentional focus instead. Many of the model properties are the same as in stage I data, but the 'positive' data is data from when attention is on a specific object, and 'negative' data is from when that object is shown on screen, but as a distractor. For each object, there are still around 15 positive examples per subject, and a maximum of 135 negative examples. At 512 Hz sample rate and 400ms analysis windows, each example is a vector of around 200 datapoints. We have 32 channels of data for both EEG and the ICA transform, so a data matrix of  $2 \times 32 \times 200$  datapoints for each of the 150 examples. We do not have to wait to compute the ICA mixing matrix, as we can reuse that from stage I. Minor technical problems with display code and recording hardware prevented some subjects from reaching this stage. We can report results from subjects 5 and 7, in 3.4.

So far, we have taken EEG data from trials in which the participant is observing a single visual stimulus. The classification accuracy may be coming from distinct EEG signatures and IC sources that are mechanistically involved with the act of viewing low-level properties of this image, or the act of attending to and perceiving this object,



some intermediate, or some combination.

By presenting a distractor image, and instructing the participant to bring her attention to the target image, we can train classifiers with the presence and absence of attention on each image. The change in predictive power of each of the ICs, with or without attention, may be useful in assessing to what degree the activity captured by that IC is mechanistically involved with attention processes.

The key results of classification with attention are shown in Table 3.3. With use of the best models trained in Stage I, and classifying new data, it was found that trials could be correctly labelled at 0.73 AUC, even in the presence of distractor shown at the same time. Models trained solely on the presence of one target image remain classifying at high accuracy in the presence of a distractor, when attention is directed to the target object.

However, when the participant is shown this same stimulus (pixel for pixel), but the attention is directed to the 'distractor', the classification accuracy drops considerably, to 0.58 AUC. When the 'target' image is absent from the trial, the accuracy remains at around this level, at 0.61 AUC, perhaps due to the negative-label distractors stimuli correctly inducing negative trial labels.

### 3.3.5 Stage III – conscious modulation of attention-dependant components with feedback

With Stage III, those ICs that appear to have greatest change with attentional state are further investigated using an online neurofeedback protocol. This allows investigation of the degree of control that participants consciously have over the activity of these attention/perception selective ICs.

Models were trained with training data specific to attention on a certain object. In this case, 'positive' training data came from trials from Stage II in which the attention was successfully directed to a certain target object. 'Negative' training data was obtained from trials in which this object was not attended to.

Key results are shown in Table 3.4. Technical problems prevented some experimental sessions from reaching this stage. For participant 7, mean S3 EEG attention model was at chance performance of 0.5 AUC. For models trained on ICA-transformed data, the attention classifiers gave higher accuracy at 0.58 and 0.57 AUC. This is somewhat lower than previous classifiers, but above chance.

Table 3.4: Table: Mean accuracy of online Stage III attention specific classification (AUC) on each participant, using selected classifier input data

<b>S3 attention specific models</b>		
	EEG	ICA
Subject 5	0.57	0.58
Subject 7	0.5	0.57

### 3.4 Examining Online Analysis

Results from this online classification experiment suggest that online classification of visual state can be performed at high accuracy (0.63 AUC). For Stage II, is seems deliberate attention to the target object increased accuracy of these classifiers by +0.1 to +0.15 AUC.

For Stage III, training new models on the basis of attentional focus had lower accuracy, but still found 0.58 AUC from a selected IC input vector. In every case, this selected IC was an alpha-band occipital IC (as seen in Fig 3.19).

#### 3.4.1 Online optimisation

The object classification results gathered in Stage I here (0.63 AUC for IC) are slightly lower than equivalent results in Chapter 2 (0.70 AUC). That could be a result of the different processing code for speed online, or, alternatively, natural participant-participant variation.

#### 3.4.2 Attention differences

With Stage I models trained for object presence, there was perhaps surprising +0.1 to +0.15 AUC performance gain when subsequently classifying overlapping stimuli with attention directed to the target object for that trial. It was not the case that the presence of a distractor image reduced classification accuracy, unless the attention of the participant was directed to this distractor. In fact, mean accuracy was even higher than in trials where target object was present, but with no attention direction.

Additionally, using models trained on the select IC, it appeared that there was little difference between classifying trials in which target was present-but-not-attended and trial where target was completely absent.

This behaviour could be taken as weak evidence towards the activity of this IC tracking target object attention, and being *invariant* with respect to the non-attended appearance of the object in the visual field.

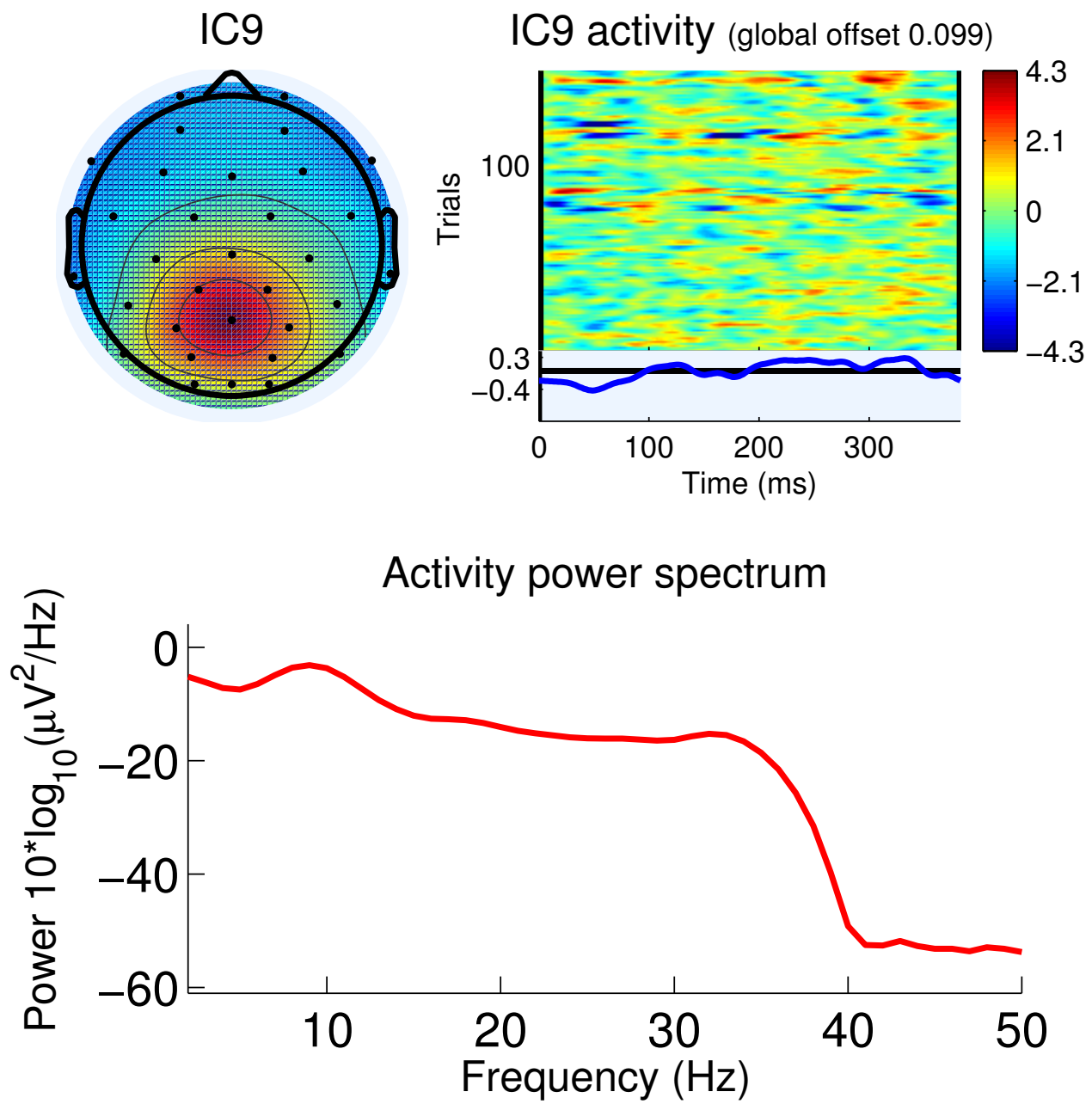


Figure 3.22: The component properties for attention-specific IC9 in participant 5, demonstrating high power at around 10 Hz

## Chapter 4

# Discussion

With the new data found in Visual Classification I and Online Analysis, we can see that support vector machines can be used to classify EEG data from visual object and visual object attention experiments. Given these results, what conclusions are supported? Are existing theories supported by this data? What lessons can be learned from this work, in terms of both theoretical support and implementation detail?

In this chapter, we address the separate areas of visual cognition, EEG processing, and machine learning, and discuss several potential conclusions from each area. With the results from the EEG classification presented, we can evaluate the core hypotheses.

### 4.1 Hypothesis 1

*Machine learning classifiers can be used to predict visual stimuli presence at above-chance accuracy.*

While machine learning classifiers had been used to good effect on EEG data for motor brain-computer interfaces (Donoghue, 2008), it was not certain that we would be able to predict visual stimuli at high accuracy, nor that our recording or analysis environment would be capable of this.

Both Visual Classification I tests of classifying individual object trials vs no-object trials, and Visual Classification I category tests of classifying a category of objects vs no-object trials show high accuracy, at 0.7 and 0.94 AUC respectively. Using the Wilcoxon signed-rank statistical test with this data against negative control data confirms this as being well above chance. It seems that this setup and protocol can indeed classify EEG data to predict presence of visual stimuli trials.

There was lower performance classifying individual object trials in a minority of subjects, but category classification had high classification accuracy in all subjects. Constructing well-suited SVM training vectors was crucial to having classifiers work correctly.

A potential refuting argument of hypothesis 1 is that these successful classifications could have occurred by chance, given the many classifiers and training data vectors

investigated, through the sheer number of comparisons we are running. That is, data could have been 'tortured until admitting it was significant'.

Multiple comparison correction and data hygiene are indeed major concerns in neuroimaging and machine learning, which can cast doubt on many results. There are several reason to say this was not such a factor here:

First, we followed good data hygiene practice. Five-fold cross validation was used for model training, reducing spurious results. Analysis code functions were blind to subject ID or any prior results, and data used for testing was kept isolated from training data. Use of AUC also reduces spurious high accuracy, as both sensitivity and specificity are required.

Second, we validate performance on this protocol by comparing trial-classifying accuracy to a 'negative control' classifier, which uses the same protocol, but with randomly offset trial start times. This negative control gives 0.5 AUC, as expected. In addition, a 'positive control', where test data is compared to trials with large blinks, gives very high accuracy, as expected.

Thirdly, in the online analysis, models are trained and specified using stage I data, before stage II takes place. These classifiers are then fed stage II trial data as test data, and give moderately high accuracy of 0.7 AUC for attentional target. Since models are trained and fixed around 20 minutes before test data is recorded, this clearly gives less room for cherry-picking or spurious data leakage.

Even with this three-pronged resistance to data hygiene and data leakage problems, it could be that they remain as a factor. However, I argue we have done enough to address this and continue.

Thus, with these results and analysis, I interpret this as being in strong support of the hypothesis that machine learning classifiers can be used to predict visual stimuli presence.

#### **4.1.1 Theoretical implications and relation to literature**

A thorough search of the surrounding literature indicates such single-trial classification performance of EEG in vision has not been reported elsewhere. There is related work decoding fMRI in vision (Andersson et al., 2011), and also profiling ERPs in visual recognition (Zhang et al., 2012), but our current experimental environment is quite distinct.

Before running these tests, we had assumed that above-chance images trial classification would probably be possible, given adjacent work in ERP, BCIs and fMRI. In retrospect, the large difference seen in ERPs in response to onset of new visual stimuli does seem salient enough that trials should be distinguishable. Large and subject-consistent visual stimuli induced ERPs were observed here, and in prior work (Vanrullen and Thorpe, 2001). Nonetheless, building the software tools, data processing, data partitioning for the classifiers, finding choice of features and data analysis were

all essential steps in this project. The experiment and results supporting hypothesis 1 have been published in Stewart et al. (2014).

While verifying hypothesis 1 was required and not trivial, our interest lies more in how these classifiers can help investigate cognition by quantifying what neural data (as observed in the EEG) can best predict cognitive state (in response to trial stimuli).

## 4.2 Hypothesis 2

*ICA can be used to separate putative EEG sources, and the resultant components can also be used with classifiers to predict visual stimuli presence at above-chance accuracy.*

We do indeed find that, by using ICA transformations, IC data can be used for visual classification in this protocol. We find peak trial-classifying accuracy from IC data, more so than any EEG channel.

One obvious interpretation of why this is the case is that ICA 'simplifies' the data through capturing maximally-independent source generators and separating them into different ICs. This is in contrast to EEG data, which has prevalent artefactual noise even after filtering.

We had assumed that IC transforms would be useful for investigating the properties of the potential sources, and were surprised to see that IC was so persistently the most trial-informative data source across so many subjects and classifier methods.

Dimensionality reduction transforms like PCA are commonly used as a pre-processing step to improve model accuracy in machine learning (Ng, 2009). It could be that ICA is acting similarly here.

### 4.2.1 Theoretical implications and relation to literature

That the peak IC was consistently the most trial-informative data source is worthy of note.

The use of ICA to isolate EEG noise is a common EEG technique, but here we show the inverse - isolating ICs of particular interest. This could be a relevant approach for many neuroimaging projects.

We can claim that the results strongly support that IC components can sensibly be used to predict visual stimuli trials.

## 4.3 Hypothesis 3

*The relative predictive performance of classifiers using the EEG channels and IC source data indicates relative relevant information in each of those sources.*

While hypothesis 3 is a pertinent research target, a strict reading could suggest that it is difficult to completely confirm or refute, as we do not have access to the 'ground truth' of what complete relevant information is.

Despite this, we have gathered results and analyses that can support different interpretations.

Of course, we are not entitled to expect our SVM to be a perfect or complete classifier. That is, there could be (and is) relevant information in the training and test data that our SVM code does not pick up. What is possible is to use reliable SVM classification as an indication that *at least some* trial-predicting information is present within these training and test examples.

Using this logic, assessing the trial-classification performance when using a specific data source can be taken as something of a lower-bound, indicating what relevant information is present in that specific data source.

Thus, we can report multiple results comparing various data sources, and show a representation of what information is contained within. It is not the maximal trial classification performance that is most interesting to us, but rather how different sources compare.

This becomes clear when examining results like those in 500ms sliding window, shown in Fig 3.6. We see trial-classification performance using 500ms of data from one IC at 0.69 AUC when using 0–500ms (where 0ms == stimuli appears on screen). The same at 25–525ms gives higher AUC, at 0.72. At later latencies, accuracy steadily drops to chance after 400–900ms. Thus, we claim there is relatively higher trial-predicting information at timepoints 25–525ms in a trial.

Tentatively, this could be interpreted as evidence of the state of the visual cognitive systems exhibiting more trial-specific activity at 25–525ms.

Additionally, we can use this trial-predicting information measure to compare across analysis conditions. In the 50ms sliding window Fig 3.7, we see classification accuracy on using only 50ms of data to train and test models. Using timepoints from 75–125ms gives peak accuracy here, at 0.64 AUC. This is larger than that of using all 500ms from 150–650ms.

Thus, we can quantify the trial-predictive information of this IC and observe that this is tightly tied to its activity at 75–125ms, and show this in light of other timepoints, sources, and analysis methods.

While comparing ERPs is a simpler way to give an indication of average EEG activity, this machine learning method has the advantage of quantifying relative trial-specific information in a more complete way. This suggests that these analyses support that this machine learning method can indicate a measure of relevant information content in an EEG example vector. We point to the caveat that not all possible relevant information will be captured by the SVM models, yet suggest the measure can still be useful.

We hope to present this method as one that other neuroimaging researchers might use to assess relative trial-predicting information content represented in different data subsets.

## 4.4 Hypothesis 4

*The properties of these sources, along with their trial-classification performance, can inform and constrain theories of how they may be involved with visual processing and cognition.*

Hypothesis 4 considers not just the trial-classification method, but how results from this might contribute to current theories and models of visual processing and cognition. As such, hypothesis 4 has deeper potential theoretical implications, and is more difficult to assess.

We can begin by looking at the properties of the newly identified results and considering how they fit with existing theories.

Dominant theories of visual recognition (for example, as discussed in 6.5 of Zhaoping (2014)) point to the fast speed of initial object recognition in the human visual system as a relevant and salient detail. Our results can contribute to this picture by presenting independent data from a novel methodology to inform precise neuroimaging activity that has trial-predicting information.

Our algorithms automatically identify a specific group of ICs that have highest trial-predicting information (see Fig 3.9), much higher than any other IC group. These occipital ICs have smooth Gaussian shape, centred on the rear of the head, and appear to capture visual alpha EEG activity with a 10Hz oscillation. It seems stimuli-locked phase-resetting desynchronisation of this 10Hz produces a highly-consistent ERP at 75–125ms.

Although identified from a different methodology, this very closely resembles the visual stimulation C1 ERP component reported in Zhang et al. (2012). This C1 component is implicated in feedforward neural activity crucial in fast vision. It is suggested that this may represent neural activity changes in human primary visual cortex (area V1) (Zhaoping, 2014).

Our algorithms can independently and automatically identify such a component, and also profile additional details of this. Since we are resolving a high-predictive ICA component in each subject, and have precise electrode location and headmaps for some subjects, we can compute approximate dipole locations of the IC source generator (see Fig 4.1). We confirm that these high-predictive ICs resolve to primary visual cortex sources in all subjects.

In this way, we can contribute to profiling the parts of theories of visual cognition mechanisms in ways that observing the C1 ERP alone cannot.

While results so far have largely been agreeing with existing elements of visual recognition theory, I suggest that this method has the strong promise to extend into identifying novel properties of human EEG in visual cognition.



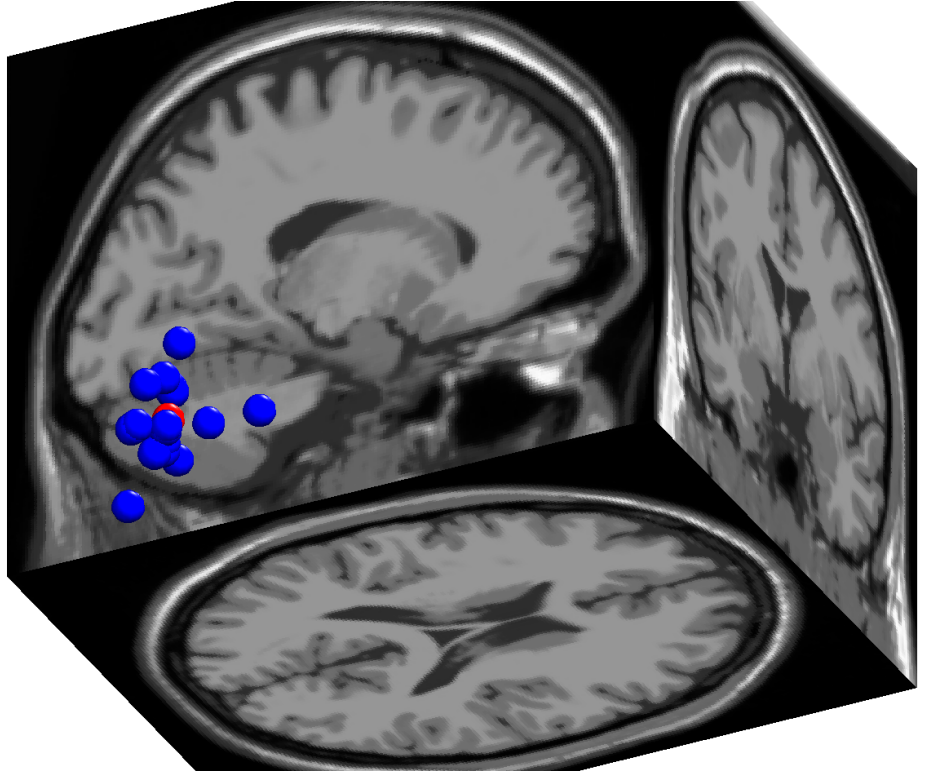


Figure 4.1: DIPFIT dipole resolution in EEGLAB, with MRI data and warped to a standard head. The selected ICs are located in V1-V2.

## 4.5 Hypothesis 5

*High-accuracy online classification of EEG data during visual presentation and target prompting can allow testing of how each source may be modulated by visual attention and intent.*

To evaluate hypothesis 5, we were able to create a new online testing environment.

Results from Stage I and Stage II show object-specific classification can be performed on single trials in this environment. While this object-specific classification task is harder than before, we also have more training examples for each object.

Stage I models of classifying object-specific trials were able to classify object appearance in the later-recorded Stage II trials at reasonable accuracy of 0.73 AUC. This was only when attention was focussed on this target image – when attention was on a distractor, the accuracy dropped to 0.58 AUC.

This can be interpreted as showing that the trial-predictive information from single-object trials is also present when a distractor image is shown overlaid, but only when attention is focussed on the target. Despite the image not changing, the act of attention being on the other object drops the accuracy.

In a planned Stage III, subjects would voluntarily modulate attentional focus, but there were limitations in proceeding to this. This was partly due to intermittent hardware issues with faulty electrodes, display code problems and Matlab crashes from

incorrectly saving an open interface. On the two subject runs where this worked well, there were tentative signs of interesting online classification feedback, but little data gathered.

As a result, I suggest hypothesis 5 is partly addressed. While object-specific and attention-specific classifiers were developed and showed some accuracy for the distractor image and increased accuracy for the attention target, we were not able to fully investigate online attention neurofeedback. The initial data obtained are insufficient to support or reject the hypothesis, but can be taken as a tentative proof-of-principle for online attentional tracking from EEG.

## **4.6 Visual Cognition**

### **4.6.1 Perception is serial in focus**

With the results in table 3.3, we can see that with models trained on object presence, accuracy increases greatly when the participants' attention is directed to said object. The accuracy does not increase in trials in which the object is present with attention directed away. This can be described as the focus of attention being the strong modulating factor, as the appearance of a distractor image does not decrease classification accuracy. When attention is directed away, even in the presence of the same pixels on the screen, accuracy falls.

This is in agreement with the focus of visual attention being a 'bottleneck', allowing serial focus on a reduced set of possible percepts. While this is fairly well established in research in the area, the new data discussed in Chapter 3 provides novel data from an original methodology also supporting this description.

I suggest that a possible implication is that an attentional bottleneck may be more fully investigated using this style of online feedback and machine learning classification from EEG data.

### **4.6.2 A single IC source located occipital-parietal region better predicts object identity**

It appears that after running ICA on EEG data, most ICs do not contain information that allows high-accuracy classification of the object presented in a trial (see section below). However, it does seem that there are several ICs that appear to offer higher accuracy. In particular, there is a single IC located in the occipital-parietal region (that is, at the rear of the head, towards the crown) that appears in most participants and gives higher trial-classification accuracy.

## 4.7 EEG processing findings

### ICA resolves relevant sources from EEG data

In agreement with previous work elsewhere (Vigario et al., 2000; Delorme et al., 2007; Hyvärinen, 2011), our data supports the hypothesis that the procedure of ICA can be used to resolve real sources of variation within EEG data.

Firstly, Infomax ICs that resembled eye blinks and muscle movement activity were present in normal trials where this behaviour was observed in the participants. When this data was used to train SVM models to classify trial state, they gave poor accuracy.

Using an IC dipole reconstruction algorithm in tandem with one participant’s fMRI data (Delorme et al., 2012), we were able to identify likely location of the IC dipole for these sources, and they were found to be in agreement with the presumed source, locating an eye blink IC to eye muscles. It also appeared that many occipital ICs had dipoles near cortical surface of visual-parietal areas.

Also, test datasets were recorded in sessions where the participant blinked heavily and deliberately repeatedly tensed jaw muscles. Running ICA on this found high activation of these IC sources.

In a verification of the experimental methodology, we found that using data from these ICs gave very high accuracy at predicting the appearance of the these blinks and muscle movements in the trials. Trials with blinks, jaw tense, or prominent alpha activity were correctly identified at 0.95-1 AUC .

While there are many recent papers confirming the use of ICA in EEG source resolution, I suggest that further evidence in favour of the utility of this tool is provided by its successful use here. Additionally, the scoring of these ICs in terms of SVM trial-predictive power using data from that IC demonstrates a way of comparing how any IC might be relevant to a given task. I suggest that reporting this alongside the spatial distribution and activity profile of an IC would allow easier comparison of EEG data mining across different groups in different labs.

### Use of 64 channels offers slightly better identification of interesting ICA sources than 32 channels

There appear to be differences in opinion as to how many electrodes are required in order to resolve the main ICA sources present (Luck, 2005). In order to resolve  $n$  ICA sources,  $> n$  electrodes are required. That is, one must have a greater number of source electrodes than the number of desired ICA sources (Hyvärinen and Oja, 2000). Many of these IC sources are likely to be unrelated to the task at hand (Hyvärinen, 2011). A valid suggestion, therefore, might be to use as many electrodes as possible; however, this is not particularly useful quantitatively.

Using data recorded here and identical processing methods and classification methods, we observed that a minority of identified ICs gave high trial-labelling accuracy classification of input data. When removing half of these electrodes from recordings

with 64 channels, and repeating the same procedure, these top few ICs were still identified in almost all participants.

With our implementation of Infomax ICA (Delorme et al., 2007), the returned ICs are typically listed in order of variance explained by each IC. In the case of either 32 or 64 channel recordings, the parietal-occipital IC that appeared related to occipital alpha rhythm was typically found in the range of the 3rd to 12th identified IC.

I suggest that our data agrees with existing advice to use 32+ channels in order to identify more possibly-relevant IC sources, while also pointing to 32 channels being sufficient to identify the ICs that gave best classification accuracy in most of our participants.

### **IC resolution is sensitive to processing, rejection and filtering procedures**

Initial attempts used EEGLAB's Infomax ICA algorithm on 64-channel EEG recordings at 2048 Hz with only minimally processed data. The practical process of running this was awkward, as it could take several days to run on a standard desktop computer and occasionally crashed Matlab.

It was found that a series of data processing steps was useful in improving data quality or practicality of analysis:

1. Re-referencing to the left mastoid electrode
2. Running a modified Hamming window FIR filter to reduce frequencies outside 2-80 Hz (or 0.5-100 Hz) (Widmann and Schröger, 2012)
3. Re-referencing again to an average reference
4. Subsampling to a datarate of either 256 Hz or 512 Hz
5. Scoring trials for noise and removing outliers (typically dropping around 1% of trials)
6. Scoring electrodes for noise levels, and removing extreme outliers (typically 1 or 2)
7. Compute ICA weights using Infomax algorithm in EEGLAB's RunICA

Following this process resulted in much smaller EEGLAB datasets of around 50MB per participant-session, rather than 500MB. Trials and electrodes with extreme noise levels are likely to be due to gross participant movement or weak electrode connection. With these removed, ICA can be thought of as having less work to do in explaining the subcomponent sources present.

I wish to stress that each of these data processing steps are mentioned in many papers analysing EEG data, but this pipeline required considerable work to optimise. For instance, use of the original EEGLAB filter was found to be inappropriate for

this data (Rousselet, 2012), potentially leading to 'smearing' of information through timepoints.

Re-referencing EEG to an average reference before computing ICA seemed to be against prevalent advice in ERP analysis (discussion in Luck (2005) recommends against it), but after discussion with Arnaud Delorme and Tim Mullen at the EEGLAB workshop in 2012 (as in Delorme et al. (2011)), recomputing an average reference was recommended before computing ICA. This was found to give faster – and perhaps more consistent – IC resolution.

For computational implementation, it was quickly identified that having a full 64 bit OS (I used Ubuntu Linux 12.04 64 bit or Windows 7 64 bit), with full 64 bit version of Matlab (where a call to *computer* returns GLNXA64 or PCWIN64), and preferably more than 8GB of RAM, allowed RunICA to run without crashing. For the offline analyses, this is easily achieved by copying the datasets to a compute server after recording, and running analysis in batch. For online analysis, a mid-spec 2012 laptop was sufficient.

## 4.8 Machine learning classification findings

### **SVM models trained on EEG data from blinks and alpha rhythm give very high accuracy**

As a validation of this methodology, I applied SVM models to single-channel data from EEG data in 200 trials in which the participant was only instructed to either blink, relax, or not instructed. The relax condition often contained strong alpha signal. Even from visual examination of the EEG trace data, it was very obvious which state each trial was from (large 100  $\mu$ V blinks, prominent 10 Hz waves, or baseline).

Our SVM models trained on this gave very high accuracy classifying subsequent test data from this experiment. It should not be surprising that SVMs can achieve high accuracy classification on clearly separable data, but failing this test would be evidence of a data handling problem, and so high accuracy here is weak evidence that our implementation of EEG processing is indeed functioning as expected.

Additionally, we can see that models trained from any single scalp EEG have high accuracy in this blink classifier. This is not the case with ICs. Many models from ICs have perfect accuracy, however models trained using some ICs have lower accuracy, despite the blink classification being comparatively easy. I suggest that this is in agreement with the activation of some ICs representing blink-related activity, and others representing EEG variation completely devoid of information about the blinks (background noise, AC noise, etc).

Importantly, the accuracy of this blink-classifier decreased progressively when I ran a negative control experiment, deliberately contaminating the SVM model training data with mislabelled trials from the same recording. This gave around 0.98 AUC on correctly labelled training data, falling to 0.77 AUC when 1/3 of the training labels were

deliberately mislabelled. This is functioning as expected, and is additional supporting evidence that the EEG processing and classification code is working as desired.

**SVM models trained on EEG data from visual object presentation trials can correctly label new trials at 0.65 AUC (85% accuracy), using only data from one selected EEG channel**

As seen in 3.1.2, visual object presentation trial classification was 0.65 AUC using RBF SVM on the selected EEG channel. The neural activity observed at the scalp between different visual objects and baseline is presumably quite subtle and complex. Thus, finding well-above chance 0.65 AUC mean accuracy across all objects and participants ( $p < 0.01$  on Wilcoxon signed rank vs negative control) was higher than I had anticipated from this lightly-processed single-channel data. Furthermore, classification of aspects of visual state from EEG at this accuracy has not been widely reported elsewhere, although classification of other modalities, such as movement intention (Hammon and Makeig, 2008), have shown promise.

While this selected EEG channel (procedure in 3.1.2) gave 0.65 AUC, the median accuracy from a single-channel was fairly low, as may be expected, as most of the EEG data might be anticipated to have only limited trial-relevant information.

**Use of IC data gives higher accuracy from SVM classifiers, at 0.7 AUC (87%) from one selected IC as input**

By applying the ICA source separation algorithm, we found that the accuracy of models trained on peak single-IC was significantly higher than those trained in the peak single-scalp-EEG, going from 0.65 to 0.70 AUC.

While the trial-classifying accuracy was slightly lower in the median IC than that of the median EEG, this agrees with a hypothesis that ICA separates independent subcomponent sources within the data, isolating those sources related to the task in a few ICs.

The significantly higher accuracy from this selected IC may then be as a result of the irrelevant noise captured in other ICs being separated from the task-informative data in this IC, and so our SVM models are more readily able to classify this at high accuracy.

**Removing the contribution of a single occipital-parietal IC (but no other IC) from datasets greatly reduces EEG trial-classification accuracy**

In almost all participants, and fully all participants with trial-classifying IC above 0.62 AUC, the selected IC with data best able to predict the trial identity was located in the central occipital-parietal region, with peak modulation around 10-13 Hz (see Fig 3.2). Not all occipitally located sources gave high accuracy, but, of all 448 identified ICs, all of the the top 5 were in the occipital-parietal region.

As shown above, trial-classification accuracy was normally around 0.65 AUC when using data from a single selected scalp electrode. When the contribution of the single selected occipital-parietal was removed from a dataset (by rejecting it from the EEGLAB dataset), and models classifying trial-classification accuracy were run again on EEG (not IC) data, peak accuracy was reduced to 0.56 AUC. Removing any other single IC did not substantially reduce accuracy.

I argue this suggests that this occipital-parietal alpha-band IC captures much of the trial-classifying information available for our models. Further, I suggest this presents this IC source as target for additional identification and probing in subsequent research.

### **Feature extraction and abstraction is key for SVM classification accuracy in complex data**

In building and applying machine learning algorithms, it is of course important that sensible training data is used for the particular problem at hand. In many problems, it is common to have multiple preprocessing steps in which input data is transformed, segmented, extracted, or filtered and only after those steps is data presented to the machine learning algorithms (as is done in face detection from SVM on 'nose segments' rather than on 'camera pixels' in (Ng, 2009)).

The aim of this process is to remove as much known-irrelevant data as possible – reducing the work that the machine learning tools need to do.

Applying SVM to raw EEG could well give some useful results in easier classification tasks, such as in classifying large blink trials versus baseline. For identifying sources related to more subtle aspects of visual cognition, this was not sufficient. We found that even after filtering, normalisation, and noisy-data rejection, the additional processing step of running ICA source separation on EEG data produced data that gave higher accuracy when used as training and test data for our SVMs.

There is also an advantage in using ICA as a preprocessing step for EEG classification, as it is 'data driven', in that the IC sources are from maximally independent subcomponents of data. In EEG data with known scalp locations, the IC sources can be spatially resolved to locations on the head. This - along with highest observed trial-classification accuracy coming from selected IC sources - suggests to me that ICA is useful in a machine learning pipeline for EEG data.

## 4.9 Potential future work

With better quantification of information on IC information flow, it may be possible to better model precise timing of steps in visual perception. We hope the greater information from EEG analyses with predictive power of each IC quantified by machine learning might give insight and constraints for top-down (or 'systems-then-lower') models of visual cognition.

This can be built on analysis of data recorded from previous experiments. Further experiments could be aimed at best taking advantage of the information that could be acquired about real time visual perception from EEG. One possibility would be to manipulate the presented visual stimuli dependant on the ongoing EEG activity, captured within a certain IC, in order to better profile that IC in time and in causal dependencies. Stimuli could be presented in a noisy overlay in which they were probabilistically reported as being consciously observed by the participant. The saliency of these images, against the noise, could be varied in real time dependant on ongoing EEG. Using our classifiers, we could monitor current visual perception. The degree to which forming percepts are 'all-or-nothing' is not well known, and this could potentially be probed within this experimental design.

If we can classify forming percepts before conscious perception, we could manipulate them. In a help-vs-hinder protocol, we could reinforce the forming percepts and compare this to attempts to sabotage the forming percept with more noise. We could then quantify how different steps are informative to this forming percept – and what information is in the ICs representing this.



## 4.10 Conclusion

While we would ideally have cellular firing patterns and LFP from every neuron in the brain, this is obviously difficult to achieve in an awake human – or even in a worm. Nonetheless, non-invasive neuroimaging does allow us to access some ways of measuring activity in behaving humans.

Work from Kirchner and Thorpe (2006) shows that object choices can be made in as little as 120ms, and that EEG data can hold information relevant to that timecourse. Detailed profiling of the ‘neural correlates’ of perceptual choices in these tests is more complex (Philiastides and Sajda, 2006).

We can build the prediction accuracy of our classifiers – using information from ICs from different spatial locations – to put additional limits on what information is where in the brain and at what latency.

Within the current project, we have demonstrated a novel approach of processing EEG data in visual experiments and training SVM models to predict the identity of the trial - stimuli or blank. With this, we were able to quantify the benefit of preprocessing techniques like ICA, identify that the majority of trial-relevant data from our SVM models was found in one occipital-parietal component and report the properties of that high-predictive component.

With online analysis, we could identify trial-predictive components as the trial was ongoing, so this data may be used as part of further experiments. With participants visual attention being prompted to one object, we can see that this occipital-parietal component gives higher accuracy object classification when attention is on the object, but accuracy falls when attention is directed away - even with the same pixels on the screen.

I suggest that this approach gives highest accuracy classification of visual stimuli from EEG that I have observed, and in combination with online analysis and attention tracking, is potentially revealing of the subprocesses of visual cognition in awake humans.

# Appendices

## Appendix A

# Appendix A - display testing

During this project, it became apparent that precise timing of stimuli display was crucial, and that this depends on the computer monitor hardware. Thus, I needed to test the monitor hardware empirically.

### A.1 Monitor testing

Older Cathode-Ray Tube (CRT) style computer monitors are typically used to present stimuli in human visual research. CRTs are capable of relatively precise timing and, often, of screen refresh rates of 120 Hz. The vast majority of new computer monitors manufactured today are flat-panel Liquid Crystal Displays (LCDs). LCDs have the advantage of emitting less electromagnetic noise than CRTs, which can be important for EEG/MEG, but some researchers have remained wary of the ability of LCD screens to present visual stimuli at precise times.

Here we test pixel transition times as a way of determining if a an LCD screen is sufficient. We present evidence that supports the premise that some – but not all – LCDs have the capability of fast visual presentation at 120 Hz. A Samsung 2233RZ 120 Hz LCD with advertised 3 ms grey-to-grey pixel transition time seems sufficiently capable for our purposes.

In psychophysics and psychological research, there is often a need to present visual stimuli to human participants with high temporal accuracy. For instance, if reaction time in response to two images were to be 20 ms different, this would be difficult to investigate if the computer screen used was slow and unpredictable in displaying these images. The most common type of computer screen used in psychophysics to date is the older, larger Cathode-Ray Tube (CRT). Since 2008, flat-panel Liquid Crystal Displays (LCDs) have outsold CRTs and are by far the most common display on new computers (Gruener, 2008), in part due to their compactness, cost and sharpness.

CRTs have the disadvantage of being expensive for large screen sizes and being bulky. A disadvantage particularly relevant for some researchers is the large electromagnetic profile of CRT that can appear as noise on MEG/EEG recordings (Wang and

Nikolic, 2011). While some advise placing the CRT monitor in a shielded room that the participant can observe through a glass window (Luck, 2005), this might not always be practical.

However, experienced psychophysics researchers have been wary of the timings in LCDs (Kleiner et al., 2007), citing slow pixel switching times and the different technology from the well-established beam flip times of CRT. Some profiles of older LCDs showed grey-to-grey pixel transition time of 15 ms to 40 ms (Artamonov, 2004), where CRTs would consistently show within 2 ms.

Recent trends toward 3D stereopsis films, TV and games have created a consumer market for fast switching computer monitors capable of 120 Hz, as 60 Hz is needed for each eye for effective stereopsis. Several newer models of LCD specifically claim 120 Hz and  $<4$  ms pixel switching speed. Here we assess a Samsung 2233RZ 120 Hz LCD monitor to test that claim.

### Monitor methods

The principle test methodology was to precisely record the change in brightness over millisecond resolution while different images were shown on the screen. To do this, a circuit was made of a fast photodiode that would respond to light intensity by passing less voltage, and do so very quickly. A photodiode similar to that shown in Farnell (2012) was used, with a rise time of  $<1\mu\text{s}$  and responsiveness around human visible light wavelength.

The oscilloscope used to record this voltage change over time was a Tektronix TDS 210. This can be triggered to record 2500 timepoints at high resolution which can be exported over serial port to a PC for recording and plotting. We used windows of 10ms-500ms.

We used the standard psychophysics research tool Psychtoolbox (Kleiner et al., 2007) to get low-level control of writing stimuli to the screens.

### Monitors used

For a CRT monitor, we examined a Sony Triniton GDM-20E01 20" CRT with maximum refresh rate of 120 Hz and resolution of  $1600 \times 1200$ , but we used  $1280 \times 768$ .

The fast LCD we examined was a Samsung 2233RZ 22" widescreen with  $1680 \times 1050$  native resolution and stated ‘response time’ of 3 ms grey-to-grey.

We also looked at a standard office LCD in a Dell 1707FP 17", manufactured circa 2006, with  $1280 \times 1024$  resolution, 60 Hz refresh rate and stated ‘response time’ of 8 ms.

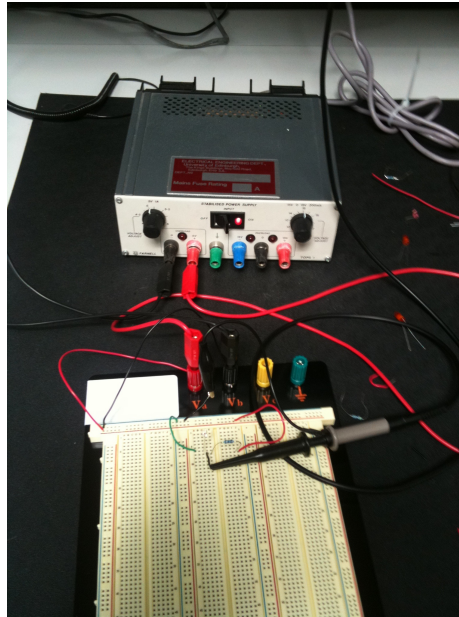


Figure A.1: The power supply (top) and breadboard with photodiode and resistor. This assembly was placed on the centre of tested computer screens.

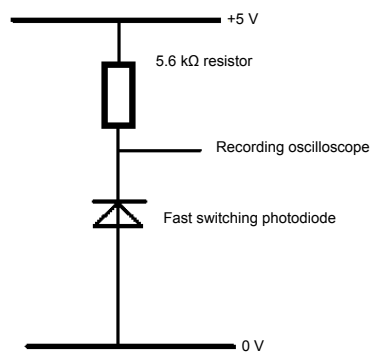


Figure A.2: The wiring diagram for the photodiode circuit. This was set up in a way to give fast and accuracy brightness readings at the oscilloscope.

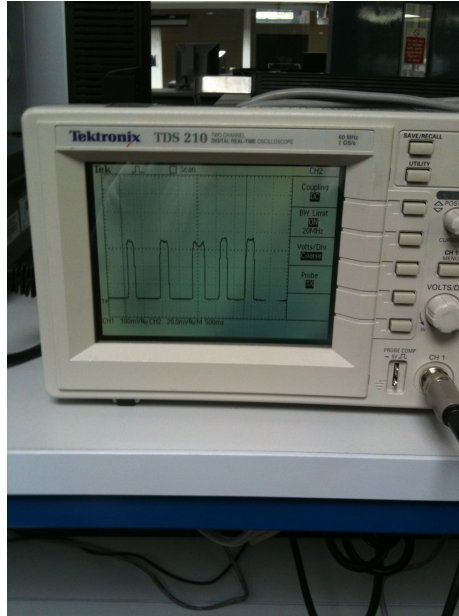


Figure A.3: The Tektronix TDS 210 oscilloscope used to record precise voltage over time.

### A.1.1 Monitor results

#### Sony CRT

The Sony CRT was first set to display a solid white screen, so recordings could be examined.

In figure A.4, we see that displaying a white screen at 120 Hz on this CRT involves 120 pulses of light, where the dark screen of 5.1 V drops to brightness of  $<4$  V within 2 ms. This brightness spike then returns to black over the next 8 ms. We take this fast brightness spike to be the electron gun of the CRT reaching the pixels next to our photodiode.

To test if this was indeed the case, the test conditions were modified. Using Psychtoolbox (Kleiner et al., 2007), a script was written to display alternating frames of saturated black and saturated white. The effect of this was that every second bright pulse was reduced to a small blip of 4.9 V brightness, consistent with the electron gun still scanning past the recording photodiode at the same time on black frames, but only weakly increasing the brightness. When the refresh rate was set as 60 Hz in software, this pulsing remained, but with the period halved. This is shown in figure figure A.5.

#### Samsung fast LCD

Using a Psychtoolbox script, a flashing screen of alternating saturated black and saturated white was displayed. The Samsung fast LCD was set to 120 Hz.

The data from a 500ms recording of the Samsung LCD is show in figure A.6. Here, there is a voltage reading of 5.1 V when the screen is black. This drops 4.25 V when

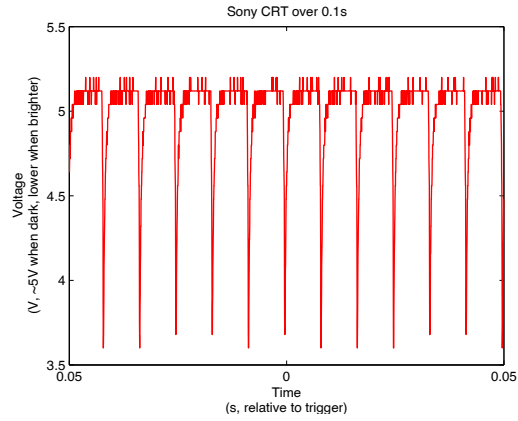


Figure A.4: A 100 ms recording of the Sony CRT displaying a solid white screen, with 120 Hz set in software. Note 5.1 V here is dark, and lower is brighter. There are 12 pulses of bright light within this 100 ms, consistent with the 120 Hz refresh rate.

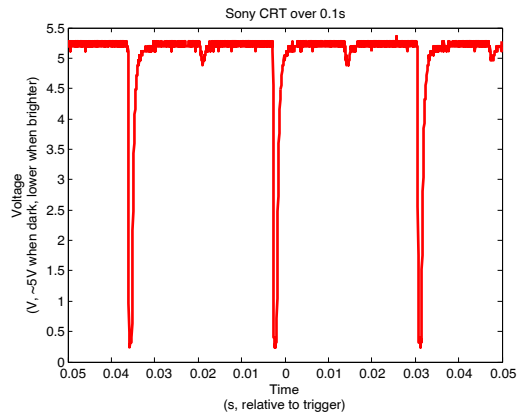


Figure A.5: A 100 ms recording of the Sony CRT displaying alternating black and white frames. For clarity here, 60 Hz is shown, but 120 Hz is the same with half period. Note 5.1 V here is dark, and lower is brighter. There are 3 pulses of bright light within this 100 ms, and 3 small pulses.

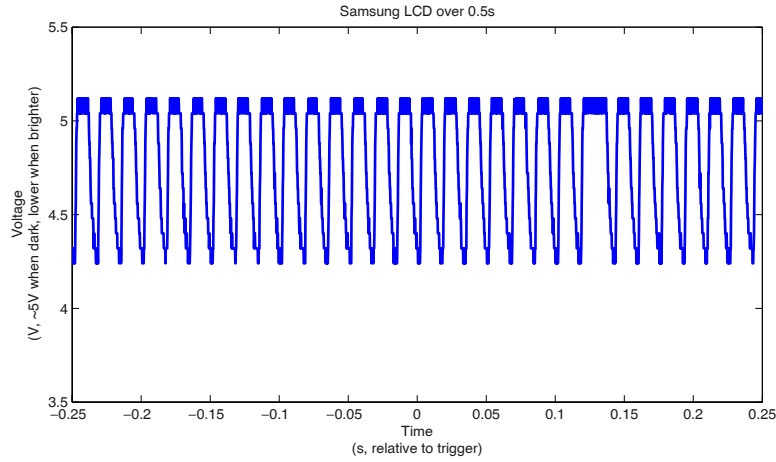


Figure A.6: The Samsung fast LCD set at 120 Hz in software, with alternating frames of black and white displayed.

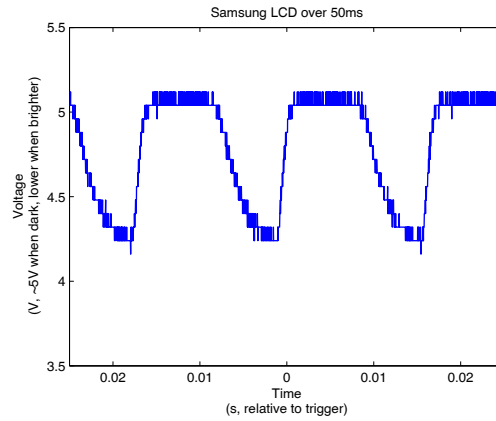


Figure A.7: The Samsung fast LCD set at 120 Hz in software, with alternating frames of black and white displayed.

the screen is white. The LCD data appears quite different from the CRT, in that there are stable periods at both dark screens (5.1 V) and bright screens (4.25 V), and not one short, large brightness spike in each frame. In this 0.5 s recording, there are 30 periods of black and 30 periods of white. This is in agreement with showing 60 frames in this 0.5 s, and so has 120 frames per second, matching the 120 Hz setting in software.

Zooming in further, data from a 50 ms recording of the Samsung LCD is shown in figure A.7. Here we see 3 periods of black and 3 periods of white, consistent with  $1/20$  of a second at 120 Hz. The black does plateau at 4.25 V. The white-to-black transition time is around 2 ms, while the black-to-white transition time is around 4.5 ms. The transition times have low variability over frames. In one display of black to white, we have 8 ms black screen, 4.5 ms black-to-white transition, 2 ms white screen, and 2 ms white-to-black transition time.

This indicates that this Samsung LCD is indeed changing pixel brightness suffi-



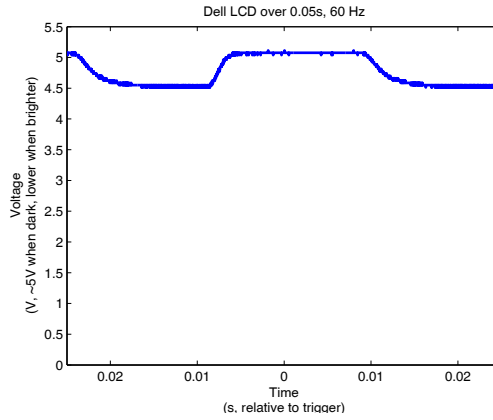


Figure A.8: A standard office Dell was also examined at 50 ms and 60 Hz refresh rate. It was set to display alternating frames of black and white. Note here black is 5.1 V and it takes 9 ms to reach maximum brightness plateau of 4.5 V.

ciently fast to keep up with a 120 Hz refresh rate.

### Standard Dell LCD

A standard office Dell LCD monitor was also examined, shown in figure A.8 at 50 ms and 60 Hz. By its design, 60 Hz was the only refresh rate that could be set in software. The black-to-white transition time of 9ms means that it cannot change pixel brightness sufficiently fast to keep up with a 120 Hz refresh rate.

### Monitor comparison

Table 1: Monitor properties

	Sony CRT	Samsung LCD	Dell LCD
Black-to-white transition time (ms)	2	5	9
White-to-black transition time (ms)	4	2	5
Capable of showing 120 Hz without loss	Yes	Yes	No

Table 1 shows a comparison of some key numbers for these three monitors. At 120 Hz refresh rate, there is 8.3 ms for display per frame. The transition times for the Sony CRT and the Samsung LCD are capable of this. The longer transition time of this office Dell LCD means that it would not be suitable for 120 Hz, and indeed that option cannot be set on this model.

Figure A.9 shows 50ms of a direct comparison of the Sony CRT at 120 Hz and the Samsung LCD at 120 Hz. The Samsung is showing alternating frame-by-frame between black (5.1 V) and white (4.25 V), while the Sony CRT is showing continuous white frames at a 120 Hz ‘clock’ for better visual clarity here. Both monitors are capable of 120 Hz.

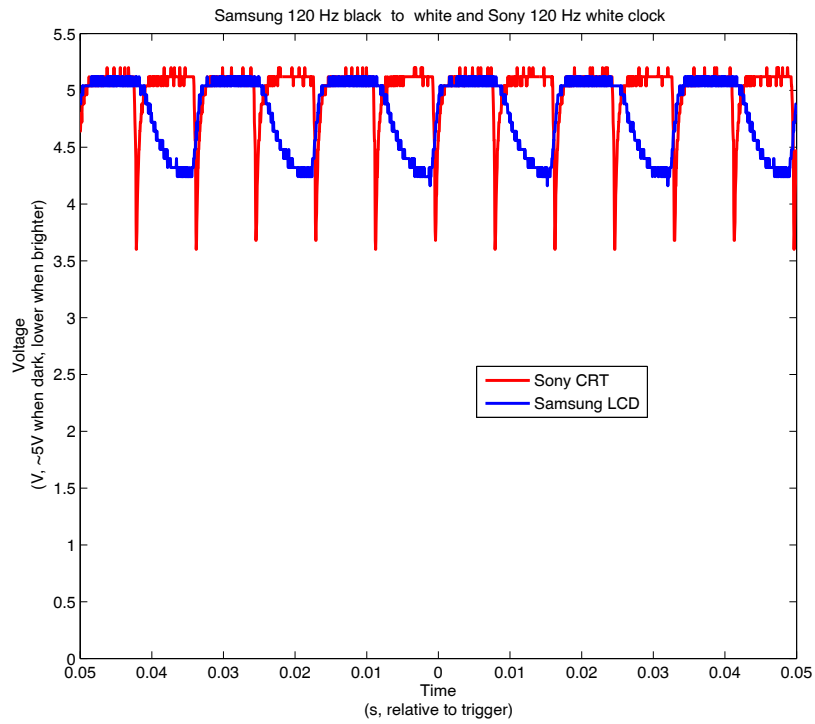


Figure A.9: The Samsung LCD alternating black and white frames at 120 Hz, shown here alongside the Sony CRT (showing continuous white frames at a 120 Hz ‘clock’ for clarity here). On each frame cycle of 8ms, both monitors are switching in less than 5 ms.

### A.1.2 Monitor discussion

Our results indicate that the Samsung 2233RZ 120 Hz LCD is indeed sufficient for fast stimuli presentation at 8 ms and 120 Hz. The pixels always changed intensity in less than 5 ms, and in a consistent manner. No shadowing or blurring was observed, either through photodiode readings or subjective use of the screen.

This agrees with findings elsewhere through independent methodology (Wang and Nikolic, 2011).

The response time of the standard 2006 Dell LCD was around 9 ms, where the manufacturer reports a response time of 8 ms. Even with this, lag and latency can be introduced by slower components feeding the framebuffer (Artamonov, 2004). The configuration of standard LCDs at 60 Hz combined with irregular grey-to-grey pixel transition times means that delays of 25-50 ms are likely.

Although the results here indicate the Samsung LCD could keep up with a CRT at 120 Hz, a limitation of the current study is that this is a relative comparison of pixel transition times. The absolute lag time between new display being requested and brightness of the screen changing is not quantified here. However, it is the pixel transition times that are the temporal bottleneck (Artamonov, 2004) and this lag time can be taken as small.

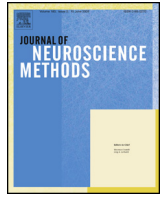
Psychtoolbox can track the precise time new screens are ‘flipped’ from the graphics card (Kleiner et al., 2007). An advantage of CRTs is that the ‘beam position’ is tracked and is a known quantity. A disadvantage is that the beam must be moved mechanically through the screen, 120 times a second at 120 Hz, to display on each pixel. That is, the pixels are painted serially. An advantage of LCDs is that the pixels can be addressed faster.

For a monitor with precise ( $<4$  ms) timing and less electromagnetic noise, a 120 Hz LCD with advertised low grey-to-grey response time seems appropriate.

## Appendix B

# Appendix B - Paper published on offline single-trials classification of EEG and ICA

Much of the material for describing EEG processing, ICA, and offline classification of trials using SVMs, appears in a paper we published in Journal of Neuroscience Methods in 2014. Here, I show this paper as published.



## Computational Neuroscience

## Single-trial classification of EEG in a visual object task using ICA and machine learning

Andrew X. Stewart<sup>a,\*</sup>, Antje Nuthmann<sup>b</sup>, Guido Sanguinetti<sup>c</sup><sup>a</sup> Neuroinformatics Doctoral Training Centre, Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, UK<sup>b</sup> Psychology Department, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, UK<sup>c</sup> Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, UK

## HIGHLIGHTS

- We consider machine learning in assessing information in different EEG data.
- We train SVM classifiers using EEG data from a visual object stimuli task.
- New data can be correctly labelled with 'object present' state well above chance.
- Using one channel of ICA data as input increases classification accuracy to 87%.
- We discuss how this method and IC sources might help studies of visual cognition.

## ARTICLE INFO

## Article history:

Received 27 November 2012

Received in revised form 23 February 2014

Accepted 24 February 2014

## Keywords:

EEG

ICA

Classification

SVM

Single-trial

## ABSTRACT

Presenting different visual object stimuli can elicit detectable changes in EEG recordings, but this is typically observed only after averaging together data from many trials and many participants.

We report results from a simple visual object recognition experiment where independent component analysis (ICA) data processing and machine learning classification were able to correctly distinguish presence of visual stimuli at around 87% (0.70 AUC,  $p < 0.0001$ ) accuracy within single trials, using data from single ICs.

Seven subjects observed a series of everyday visual object stimuli while EEG was recorded. The task was to indicate whether or not they recognised each object as familiar to them. EEG or IC data from a subset of initial object presentations was used to train support vector machine (SVM) classifiers, which then generated a label for subsequent data. Task-label classifier accuracy gives a proxy measure of task-related information present in the data used to train.

This allows comparison of EEG data processing techniques – here, we found selected single ICs that give higher performance than when classifying from any single scalp EEG channel (0.70 AUC vs 0.65 AUC,  $p < 0.0001$ ). Most of these single selected ICs were found in occipital regions. Scoring a sliding analysis window moving through the time-points of the trial revealed that peak accuracy is when using data from +75 to +125 ms relative to the object appearing on screen. We discuss the use of such classification and potential cognitive implications of differential accuracy on IC activations.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

## 1. Introduction

Electroencephalography (EEG) allows neuroimaging with high temporal resolution. This can be used for investigation of the properties of cognitive neuroscience processes, like the speed of

early visual processing in humans. For example, EEG Event-Related Potential (ERP) differences can be detected in animal versus non-animal images within 150 ms of presentation (Thorpe et al., 1996) and faces versus shapes within 85 ms (Mouchetant-Rostaing and Giard, 2000).

However, ERP averaged over many subjects and many trials is not the totality of the information present in EEG recordings. The use of ERP alone has been criticised (Rousselet and Pernet, 2011; Gaspar et al., 2011) for masking effects within grand averaging and not using all available information.

\* Corresponding author at: 2.53 Informatics Forum, Edinburgh EH8 9AB, UK.

E-mail address: [andrew.x@ed.ac.uk](mailto:andrew.x@ed.ac.uk) (A.X. Stewart).

URL: <http://www.anc.ed.ac.uk/dtc/find.php?andrewstewart> (A.X. Stewart).

Advanced data mining of EEG data (Makeig et al., 2004) has been suggested as an alternative approach, taking into account other aspects, such as spectral power, phase, trial-to-trial consistency and ICA activations. It may be more difficult to interpret information integrated from these multiple aspects than it is to observe grand-averaged ERP plots. How to compare these techniques is then a relevant question.

### 1.1. Machine learning with EEG data

One approach to assess the use of these techniques is to use machine learning on EEG data. Training examples from different experimental states can be used to train machine learning classifiers, and the ability to correctly and robustly identify new examples can be assessed.

This is done in the adjacent field of brain–computer interfacing, where EEG signals are used to control motor prostheses (Wolpaw et al., 2000; Donoghue, 2008). For EEG motor prostheses, the experimental states of interest might be presence or absence of prompted hand movement and success of the classifier would be to robustly identify this motor movement in subsequent EEG data.

We hypothesise a similar approach could be applied to EEG data recorded in visual object presentations. Instead of identifying motor movement such as ‘grab’ from resting baseline with no movement, the experimental states could be the presence of a given visual object stimulus (e.g. ‘spoon’) and resting baseline with no visual stimulus.

In EEG motor prostheses, there is a desire for accurate and fast classification. Consequently, many analysis methods have been attempted (Sajda et al., 2003; Lotte et al., 2007). Use of data pre-processing and machine learning tools have proved effective in improving the ability to use EEG data to predict movement intent (Müller et al., 2008).

The use of machine learning on visual EEG experiments is not as well studied as EEG data in motor movement. The EEG response elicited from different object stimuli is likely less regular, more subtle and likely has a lower signal-to-noise ratio than EEG from different motor movements, but we suggest recent advances from motor brain–computer interfaces might also apply to improving learning from EEG data in visual experiments.

Therefore, we designed a simple visual object presentation experiment where we labelled EEG data (as ‘object present’ or ‘object absent’), trained an SVM classifier with this data and labels, and assessed that model’s accuracy at labelling subsequent unseen EEG data correctly. We used classifier task-labelling accuracy as a metric of task-relevant information in the data used to train the classifier. In the present context, this allows comparison of EEG data processing methods.

### 1.2. EEG features and ICA

In the field of machine learning, the input channels of the data that are used to train classifiers are termed ‘features’. To improve the accuracy of brain–computer interfaces, a useful technique has been to focus on a small spatial subset of relevant features from EEG. An example of this is identifying ‘common spatial features’ (Müller et al., 2008; Wang and Jung, 2012). It is well known that these motor signals are highly localised in the motor strip and pre-motor areas. This made identification of reduced subsets of relevant features tractable (Blankertz et al., 2008). A comparable feature extraction procedure has not yet been accomplished for visual EEG data.

In order to give a comparable decomposition that could be relevant in vision, independent component analysis (ICA) was used on the EEG data in the present study. Each of the IC sources found from the EEG can be considered a reduced subset of the

activity in the EEG. In essence, it separates the main independent generators of variance within a signal (Bell and Sejnowski, 1995). An illustrative example of ICA is the ‘cocktail-party problem’ where, given data from three microphones in a room and three overlapping voices, ICA can separate the three individual voices as three distinct sources (Hyvärinen and Oja, 2000). For EEG data then, the detectors of EEG electrodes act as the microphones and the varying electrical patterns as the voices.

ICA is typically used for identifying and removing noisy electrodes, blinks and other artefacts to ‘clean up’ EEG data (Luck, 2005) before proceeding with conventional ERP analysis. Here, we instead use ICA as another way of describing the EEG data, to give subsets of data that may be both more interpretable and give higher classifier performance. The ICA components provide an estimation of possible ‘sources’ of generated activity. This can be particularly advantageous in EEG data analysis, where much of the signal (and noise) is shared across all channels (Onton et al., 2006). In summary, some of the identified IC sources may be useful descriptions of subsets of variation within the EEG, as ‘common spatial features’ have been in motor EEG.

### 1.3. Support vector machines classifiers

We sought to determine whether EEG data in our experiment can be automatically classified using machine learning tools. To this end, we used support vector machines (SVMs) – a flexible and powerful statistical learning tool (Burges, 1998; Cortes and Vapnik, 1995). This technique has given particularly good results in a wide range of domains, including cancer classification (Furey et al., 2000) and face detection (Osuna et al., 1997). In classification of EEG, SVMs have shown good performance in many contexts (Lotte et al., 2007). SVMs were the most commonly used technique for highest accuracy in an EEG classification competition (Blankertz et al., 2006). Thus, we used SVMs as our classifier here.

Specifically, we used SVMs to classify the EEG data into two classes according to the presence or absence of specific visual stimuli. The underlying principle of SVM classification is to solve a (non-linear) classification problem by transforming it into a linear classification problem in a different, higher dimensional space (or *feature space*).

This is achieved by introducing a non-linear map (*feature map*) into the feature space, which can often be an infinite dimensional space of function. The important aspect of this procedure is that, for many widely used algorithms, one is only interested in the scalar products between pairs of feature vectors; these scalar products can be computed by means of a *kernel function* which depends only on the original (non-transformed) data points. Therefore, the need to work explicitly in the high dimensional space is removed and the feature map is defined implicitly by a choice of kernel function.

A commonly adopted kernel is the so-called *Radial Basis Function* (RBF) kernel; the scalar product of the images of two data points  $x$  and  $y$  under the feature map implied by the RBF kernel is computed as:

$$k(x, y) = \exp[-\gamma \|x - y\|^2] \quad (1)$$

where  $\gamma$  is a tunable parameter.

Once a kernel function is selected, the SVM algorithm works by identifying a hyperplane in feature space that optimally separates the two classes in the training data, giving the maximum margin between the images in feature space of the points in the two classes. Often it is desirable to allow a few misclassifications in order to achieve a wider margin of separation; this trade-off is controlled by another parameter, called the training error cost, and usually denoted by  $C$ .

#### 1.4. Classification of visual object presence from EEG data

Previous studies of visual perception using EEG have demonstrated significant task-specific ERP differences within 150 ms of stimuli appearing (Thorpe et al., 1996). In their study, human subjects were able to perform a rapid decision on whether or not a visual scene contained an animal. Notably, examination of the EEG data was able to show the evolution of the difference in ERP on animal and non-animal trials.

Researchers interested in visual perception might ask what the source of such ERP differences is. To what extent can EEG signatures be used to decode observed images? What parts of the EEG signatures are useful for this?

The current study aims to extend this examination of EEG differences upon the presentation of single images by quantifying classification performance from different transforms of the EEG data, and EEG data from different times.

In designing a classifier of visual stimuli given EEG data, we first sought to determine if a visual object presentation could be distinguished at all. EEG data recorded in each visual object presentation trial was used along with more EEG data from resting baseline (blank screen), to train a machine learning SVM classifier. This was used to predict the correct label (as object present or absent) of additional unseen EEG data.

If the EEG data used to train the classifier contains information that is indicative of one state or another, it might allow high trial classification accuracy. We can choose what parts of the data are used for training and classifying, and so search for those parts of the EEG data that are most relevant for distinguishing trials.

As ICA gives an estimation of sources within the signal, we can use this classification procedure to assess the trial-classifying performance of each of these IC sources.

#### 1.5. Hypothesis

Here we propose a method for classifying visual object presentation from EEG readings using machine learning. We test whether, using SVMs, we can classify the presence of visual object stimuli at above chance levels.

Further, we wish to compare scalp EEG data to EEG data transformed into ICs. Since ICA is a data reduction technique, the expectation is that some ICs might give higher accuracy.

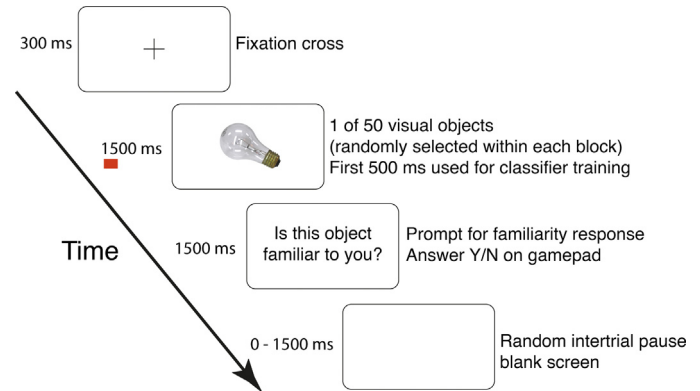
## 2. Methods

#### 2.1. Participants

Seven participants (five women, two men; median age of 25) took part in the study and were each compensated £7 per hour. Ethical approval was granted by the University of Edinburgh Psychology Research Ethics Committee. Participants self-reported normal or corrected-to-normal vision and gave informed written consent.

#### 2.2. Materials and design

Participants were instructed to observe a series of object images presented to them on a computer screen. Fifty colour photographs of common real-world objects were selected from the Bank of Standardized Stimuli (BOSS) provided by Brodeur et al., 2010. Multiple presentation trials of each object were required for classifier training and testing. There were five blocks of 50 trials, where each object was shown once in a given block, in random order.



**Fig. 1.** Experimental design of the visual object presentation protocol. The time-course of the events is shown. Participants were shown a fixation cross before stimulus presentation of an everyday visual object – in this example, a light-bulb – compiled from a published standardised image bank (Brodeur et al., 2010). There were 250 of these trials – five presentations each of the 50 stimuli, ordered randomly. The red bar indicates the 500 ms after presentation that was used as object onset data to train classifiers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

#### 2.3. Apparatus

The experimental script was written in MATLAB 2009b (Mathworks, Inc., Natick, MA, USA). Extensions from Psychophysics Toolbox 3 were used for better timing precision (Brainard, 1997; Kleiner et al., 2007). Stimuli were displayed on a fast-refresh 22 inch Samsung SM2233RZ monitor at 1000 × 1000 pixels centred on a 1680 × 1050 pixel display, with the participants leaning on a chin-rest 72 cm away, giving an object size of 15–22° in the visual field. The complete screen was around 38° by 24°.

#### 2.4. Experimental procedure

Each trial began with a central fixation cross for 300 ms, followed by a randomly selected object image that was presented for 1.5 s (Fig. 1). An on-screen text prompt then asked the subject if they had recognised this object as something familiar to them. This was implemented to avoid passive viewing of the stimuli. Subjects responded using buttons on a gamepad – one marked ‘yes’ if this object was recognised as familiar and one marked ‘no’ otherwise. The text was replaced by a blank screen on button response or after 1.5 s. There was then a random intertrial interval of 0–1.5 s, where the screen remained blank until the next trial began. The experiment was performed in a single session per subject of around 1.5 h, in which stimuli presentation and breaks lasted around 25 min.

#### 2.5. EEG recording

EEG was recorded from 64 head electrodes and six support electrodes using a Biosemi ActiveTwo amplifier at a sample rate of 1024 Hz. These 64 electrodes were placed according to the standard 10/20 EEG electrode system, and held in place using a Biosemi electrode cap of appropriate size for the participant. Triggers generated from the experimental code were recorded on the EEG device to allow timing synchronisation. All electrode offsets were below 20 mV. The six support electrodes were placed on two mastoids, two temples, and above and below the right eye, for better EOG detection. The experiment took place in a shielded experimental room.



## 2.6. EEG data processing and ICA

The processing of the raw EEG data was performed using custom code that included use of standard EEG processing functions from the EEGLAB v12 toolbox (Delorme and Makeig, 2004) in MATLAB. Biosemi data was loaded using left mastoid reference, and re-referenced to an average reference later (Luck, 2005). A Hamming-windowed FIR band-pass filter of 0.1–80 Hz was applied, using 'eegfiltnew' in EEGLAB v12. The 1024 Hz recording was down-sampled to a sample rate of 256 Hz using EEGLAB functions.

Data rejection was performed using standard EEGLAB functions. Noise of four times that of the median was chosen as a threshold. Electrodes with values exceeding this criterion were removed from subsequent analysis. This resulted in a median of 20 channels rejected (range: 11–22 for the seven subjects), leaving a median of 49 electrodes. Most of the head remained well represented as no large areas were left without any electrode coverage.

Independent component analysis (ICA) was applied to the whole dataset from each experimental session, after artefact channel rejection, using the Infomax ICA algorithm (Delorme et al., 2007). Infomax ICA here returns one IC for each electrode, giving a median of 49 ICs for each subject. The timing triggers were used to label the time-points in which each image was presented.

After rejection and ICA, but before classifier training, each data vector was normalised by subtracting the mean of each channel and dividing data by its standard deviation. This scales the data for the SVM classifiers.

## 2.7. Classifier training

Separate classifiers were used for each subject. In the training phase, a 'one-versus-one' SVM model was trained for each of the 50 objects presented, where the labels were 'object appears on screen' or 'no object present'. The classification task was to best apply this label appropriately to subsequent data, given this training data. That is, rather than using a single multi-class classifier that would give output of 1 of 50 labels, we used 50 binary classifiers each labelling 'object stimuli present' or 'object absent' for their respective object stimuli. Both the voltage time-points from EEG and the independent component (IC) transform activations were normalised (Fig. 2). The kernel used with the SVM was a radial basis function (Keerthi and Lin, 2003) from the Matlab implementation of libSVM (Chang and Lin, 2011, software available at the libSVM website).

To find appropriate values for the SVM hard margin training error cost 'C' and the radial basis function kernel parameter ' $\gamma$ ', we used the libSVM parameter sweep tool. A grid search was performed on 18 parameter values between  $C = [10^{-2} \text{ to } 10^{10}]$  and  $\gamma = [10^{-4} \text{ to } 1]$  on data from two subjects. This suggested values of  $C = 1$  and  $\gamma = 1/\text{number features}$ , and those parameter values were used for all models.

## 2.8. Classifier input

Each object had five presentation trials. Classifiers were trained using four of these and tested using their accuracy at predicting the appearance of each object on the fifth trial.

The training data was EEG data beginning at the time of object presentation until some time after (initially 500 ms). For each object, a 'positive training data' label was applied to data from trials where that object was shown. Resting baseline EEG data, in which no visual stimuli was displayed (blank screen), was termed 'negative training data', where the correct label is 'object absent'. These training data were labelled as '1' and '0' respectively when used with our classifier.

For each object classifier, positive 'object present' input data came from the 0.5 s after each presentation of that object. As we used four trials, with each trial giving 0.5 s of data in which an object is present, we acquired 2 s of image-present data for each object. As a sample rate of 256 Hz was used, this gave a positive example training vector of 512 data points  $\times$  number of channels.

For the negative data, 'object absent' input data came from 19.5 s of randomly selected intertrial data. Classifiers were trained using a proportion of approximately 10% 'object present' data and 90% 'object absent' data – 2 s of data to 19.5 s of data. Test data also used this approximate proportion of data.

Test data came from the one presentation trial of each object that was not used for training (rotated as explained below) along with further negative examples of random intertrial data in which no object was presented.

## 2.9. Cross-validation

To improve both the robustness of the classifiers and their ability to generalise to new data, five-fold (leave-one-out) cross-validation was used (Efron and Gong, 1983). This process reduces the likelihood of erroneous results, as multiple splits of the data are considered. Cross-validation was performed on data from each object by dividing the data into five splits, with a single object presentation trial in each split. We then iterated through five separate SVM models – each training on four of five trials. The remaining fifth trial was then used as a blind test. As we iterated through the cross-validation, each trial was used once as a test data.

Classification test results then came from the accuracy on classifying the respective unseen fifth, and the score was averaged across the five splits. All classification results reported are from this average of all five cross-validation splits.

## 2.10. Performance metrics

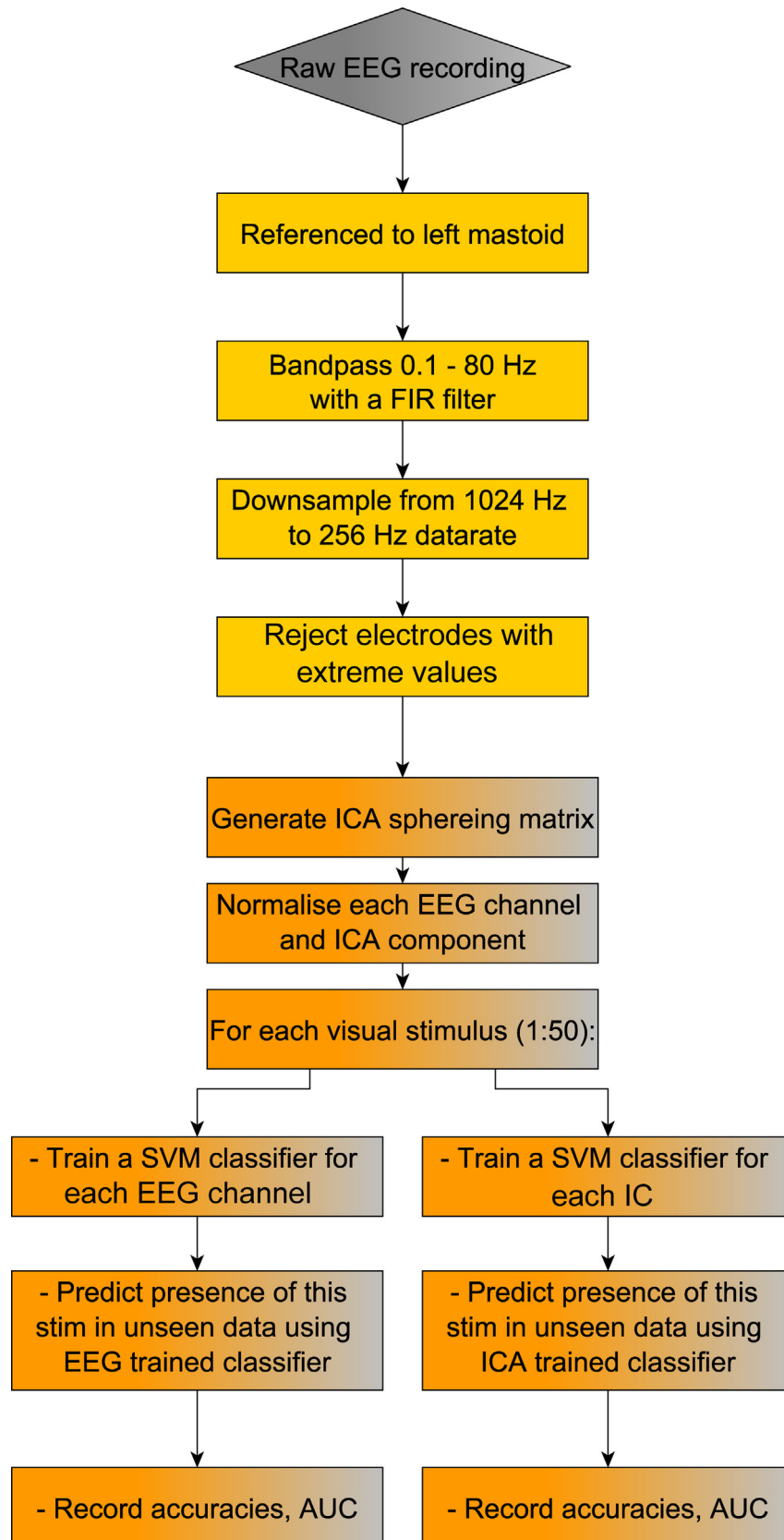
The success of a classifier can be given simply as percentage correctly classified. This can be valid in many contexts, but does not clearly show that performance depends on both 'sensitivity' (true positives) and 'specificity' (true negatives). A receiver–operator characteristic (ROC) plot (Mason and Graham, 2002; Hand, 2009) illustrates both sensitivity and specificity – with the area under the curve (AUC) of the ROC of 0.5 signifying random chance prediction and one being perfect prediction. This was relevant here as over 90% of data in both training and test situations belonged to the 'object absent' class. If a classifier were to predict 'object absent' everywhere, it might get 90% accuracy despite conveying no useful information. AUC, however, would correctly score that classification as no-better-than-chance performance.

Plotting a ROC curve can be particularly useful when sensitivity and specificity are being manipulated separately, but here we simply use area under this curve (AUC) as a concise metric of both classifier sensitivity and specificity.

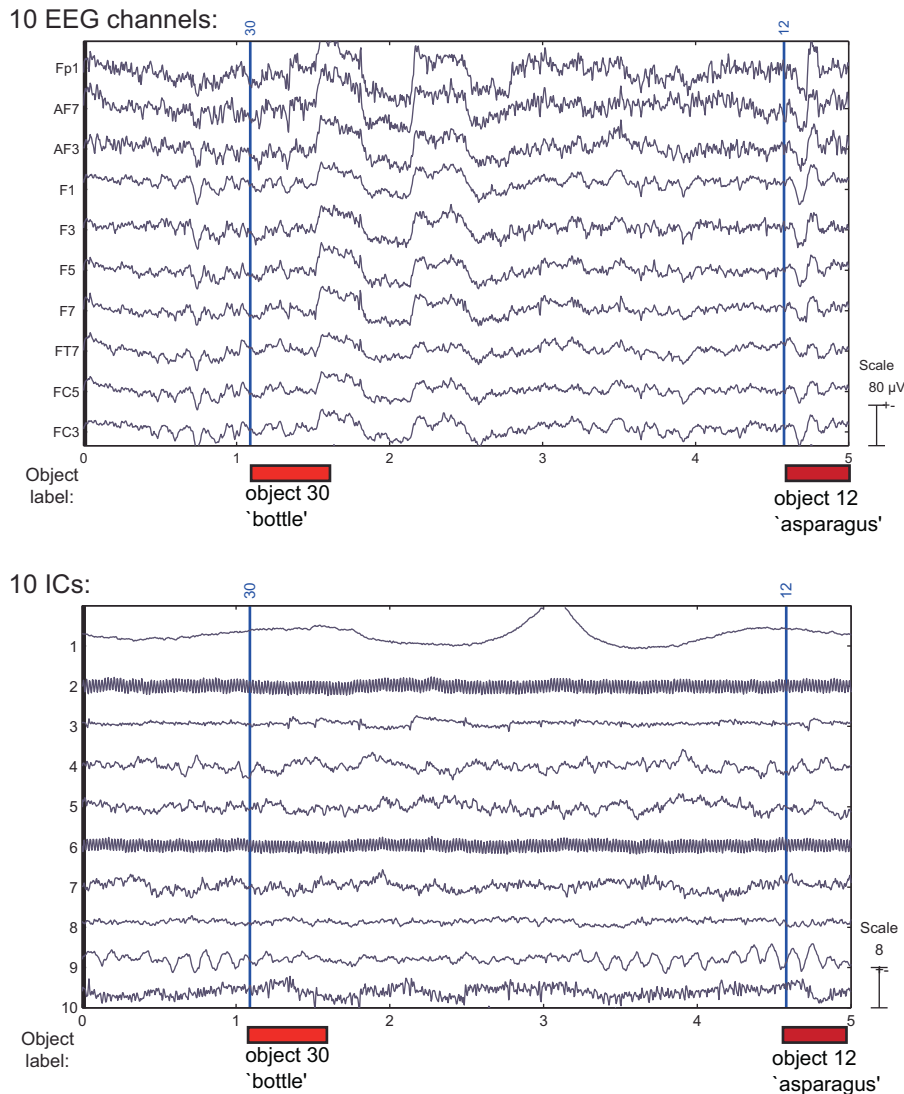
## 2.11. ERP waveforms

After data processing, we confirmed that our visual stimuli presentation elicited a visually evoked potential change. EEG trace data was visually inspected (Fig. 3) to check recording quality for EEG artefacts. Fig. 3 (top) shows EEG trace from one trial on 10 EEG channels. No obvious EEG artefacts are present. The thin blue line around 1.1 and 4.6 s into the recording indicates presentation times of two objects, with the times labelled indicated with the red bars.





**Fig. 2.** Data flow for EEG processing. Data processing steps are shown in yellow and machine learning steps in orange. The input here was the scalp EEG recording, which was referenced to the 'mastoid' electrodes behind the left ear. A FIR filter was used to reduce signal outside 0.1–80 Hz. The sample rate was then reduced from 1024 to 256 Hz to speed subsequent analysis, and electrodes with very high electrodes noise were rejected. The ICA was run and SVM classifiers were trained for all objects using EEG channels data (left) and IC activation data (right).



**Fig. 3.** EEG trace and IC activations in an example trial. The top panel displays example EEG trace data from 10 channels for 5 s around the first visual object presentation (subject 1, trial 1). The lower panel shows the data from the same period in data from 10 ICs. The blue vertical lines show object presentation time (object 30, 'bottle', shown at around 1 s and object 12, 'asparagus' shown at 4.6 s) and the red bars indicates the approximate time labelled as 'object present' used for training classifiers. The intertrial time was labelled as 'resting baseline' (no object present). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3. Results

We present results from classifying new EEG as either 'object present' or 'object absent', based on SVM models built on prior EEG data. We first report results using 500 ms of data and from classifiers using all EEG data, classifiers using each of the individual channels of EEG data, and classifiers where only one selected channel is used. We repeat this with IC data.

In addition, we report details and dynamics of high-accuracy ICs. Furthermore, we consider classification from EEG channels in which artefactual ICs are removed. Timecourse analysis is examined in classification on single EEG channels and IC data in 500 and 50 ms 'sliding windows'.

#### 3.1. Classification of visual trials from EEG data

Training SVM classifiers based on a data vector of all channels simultaneously (median number of 49 channels) appeared to give low accuracy (around 0.51 AUC). When training classifiers on data

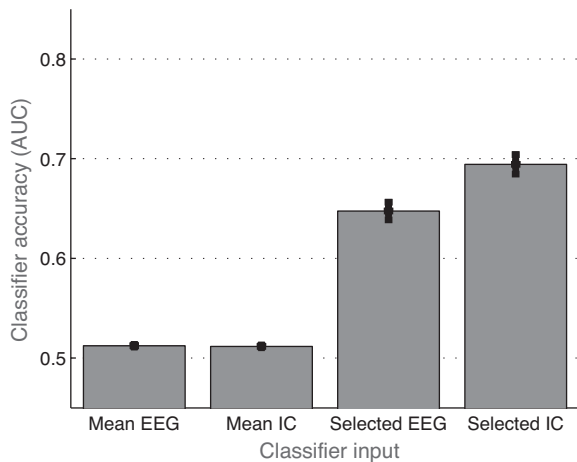
vectors from a single individual EEG channel, mean accuracy was found to be 0.51 AUC (see Fig. 4, leftmost bar).

#### 3.2. Finding a single selected EEG channel/IC

The analyses here considers EEG channels and ICs that gave the highest accuracy. A small subset of channels consistently gave much higher classification accuracy.

To objectively identify these, we used an automated selection procedure, based on data separate from the test data. These high accuracy channels can be selected automatically by further partitioning the data (prior to testing) in a training and validation set. In this way, the generalisation ability of each individual channel can be assessed on the validation set, and then confirmed (or not) on the test set. We refer to these as 'selected channels'.

Specifically, data from 10 of the 50 objects was separated. On this data, classifiers were trained using data from each individual channel. That is, around 49 separate classifiers were trained for each object, using data from each of the data channels (median number of 49). The single data channel that gave highest AUC on these



**Fig. 4.** Mean classifier performance across all subjects and objects using EEG (mean of all channels), ICA (mean of all channels), and the selected single channel for EEG and ICA. Classification performance is measured in successful classifications area under a ROC curve. The error bars show the standard error of the mean. With 50 objects shown five times each to the seven subjects, and models trained and tested for each scalp electrodes or IC, the mean is an average of thousands of classifications. The 'selected' EEG channel or IC was a single channel selected on the basis of performance on separate data, with the selected IC here giving 0.7 AUC.

first 10 objects was then used as the single 'selected' channel. The selected channel was used as the best input to classifiers for training SVMs on the remaining 40 objects. Note that this data selection procedure was performed once on each participant for both EEG and IC data (see below).

In the case of using scalp EEG data, this selection of one channel input increased classifier performance from 0.51 AUC mean to 0.65 AUC (Fig. 4, bars 1 and 3). We show the results of selected EEG channel for each subject in Table 1. Percentage of trials correctly classified is shown in Fig. 5.

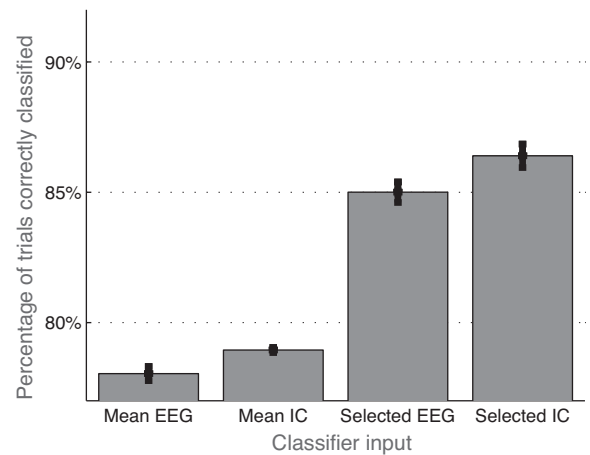
### 3.3. Classification of visual trials using ICA data

We wished to assess the classification performance on EEG data when transformed with ICA into single ICs. This IC activation data was acquired by running ICA on each participant's entire EEG recording.

As with the mean of all scalp EEG channels, the mean of ICs was low, around 0.51 AUC. Repeating the same 'selection procedure' as in EEG data above, we found single ICs gave greatly increased accuracy of 0.70 AUC (see Fig. 4).

Classification accuracy was significantly higher when using the selected IC as opposed to selected scalp EEG channel using a Wilcoxon signed rank test ( $p < 0.0001$ , means 0.70 AUC and 0.65 AUC).

These results can also be summarised in terms of mean percentage of trials correctly classified (Fig. 5). We find that these agree



**Fig. 5.** Mean classifier performance in terms of percentage of test single-trials correctly labelled. The error bars show the standard error of the mean. The selected IC gave 87% of single-trials classified correctly, across all subjects and objects, which corresponds to 0.7 AUC.

with the AUC reported in Fig. 4, with the selected IC giving highest raw average classifier performance of 87% trials correctly classified.

The above results demonstrate high accuracy (0.70 AUC, 87%) at correctly labelling 500 ms of EEG data from single ICs with 'object present' or 'object absent'. We can now consider the properties of these high accuracy ICs, examine their accuracy over time, and why they might offer higher accuracy.

### 3.4. Properties of the selected EEG channels

As shown in Fig. 6, the selected EEG channel was different in each subject, although mostly occipital (O1, O2) or parietal (P3, P5) electrodes. On presentation of visual object stimuli, we found the mean grand-average ERP (Fig. 6) with a prominent positive voltage deflection around 100–200 ms after presentation on central and parietal electrodes. This is in agreement with ERPs observed elsewhere – e.g. visual object presentation in Rousselet et al., 2007.

### 3.5. Properties of the selected ICs

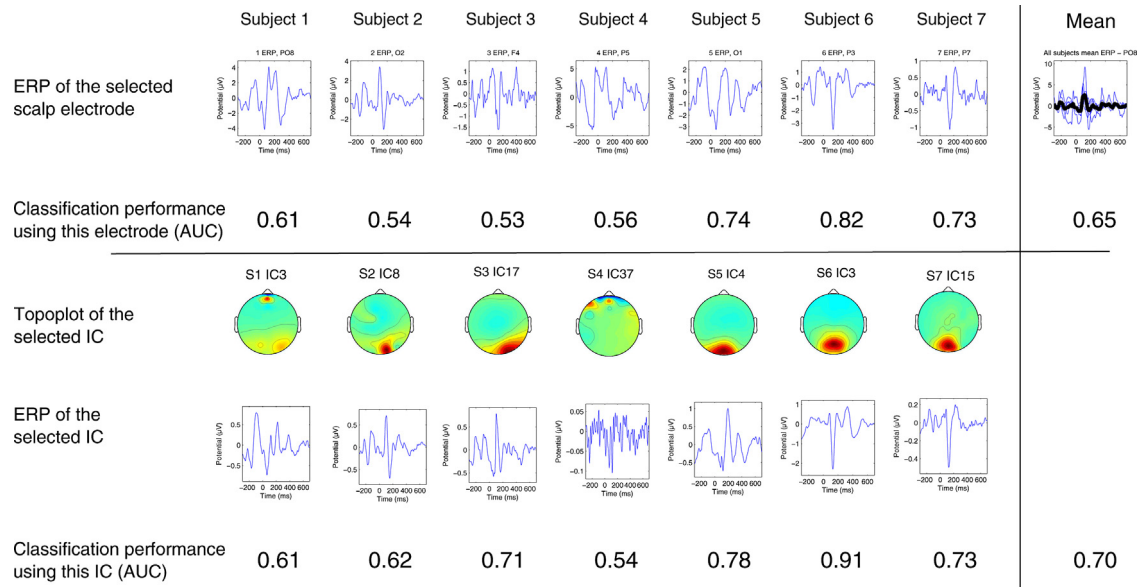
Component activations of the selected ICs in each subject are shown in the lower section of Fig. 6, along with accuracy of each IC. A 'topoplot' shows the spatial distribution of each IC on a 2D headmap. The majority (5/7) of the ICs appear concentrated in occipital regions. All those ICs with accuracy of above 0.62 AUC were found to be in this region.

Extended details of the properties of a selected IC and a low predictive power component are compared in Fig. 7. The IC activations were mapped to spatial locations on the head (Delorme and Makeig, 2004). This mapping is shown on a standard 3D head model, for both a high and a low predictive IC. Also shown is the average ERP and the trial-by-trial 'ERP image' heat-map. The ERP image is a coloured plot in which every horizontal line shows activity from a single trial (Makeig et al., 2004). In this way, the variation in individual trials can be shown alongside the averaged ERP. Here, IC 3 from subject 6 is the 'high predictive' IC, and has a complex activation pattern over time, but seems to have a 12 Hz oscillation that flattens around 100 ms after object presentation. In contrast, the activation of the 'low predictive' IC 14 is not consistent across trials and does not have consistent timing in relation to the stimulus appearing, and appears to be an irrelevant frontal IC source.

**Table 1**

Accuracy of visual object classification (AUC) on each subject, using selected classifier input data.

Input	EEG selection (AUC)	IC selection
Subject 1	0.61	0.61
Subject 2	0.54	0.62
Subject 3	0.53	0.71
Subject 4	0.56	0.54
Subject 5	0.74	0.78
Subject 6	0.82	0.91
Subject 7	0.73	0.68
Mean	0.65	0.70



**Fig. 6.** **Top** Scalp EEG ERPs from the selected electrodes in each of the seven subjects, along with the object classification performance when using data from that electrode. Thus, electrode P08 is the selected electrode in subject 1, and when data from P08 is used as input for SVM classifiers, object presence is correctly identified at 0.61 AUC. In subject 6, P3 is found to be the better electrode and it gives 0.82 AUC in that subject. **Bottom** The lower section shows the selected IC for each subject. The 2D headplot (using EEGLAB's 'topoplot') shows the spatial distribution of each of the selected ICs. Note that all ICs above 0.62 AUC (5 of 7) appear in the occipital region. The IC ERP dynamics are shown to indicate the mean activity of a given IC across all trials. The figure also displays the classification performance from Table 1, with the selected IC in subject 6 having the highest AUC (0.91).

### 3.6. Scalp EEG data with IC artefacts removed

The above results suggest improved classification when using data from activations of components identified by ICA with the SVM classifiers rather than using scalp EEG data, at least when a selected higher accuracy data channel is selected. A possible explanation for this might be that relevant neural signals are overwhelmed by electrical artefacts that are not relevant to the visual state and these are present on all electrodes. Then, when ICA is run, the contributions of those artefactual signals are 'captured' in single ICs. In order to assess this, we now remove the contribution of each of those probable artefactual ICs from the scalp EEG data and test the visual state prediction again. This effectively removes the contribution of these ICs from the EEG data. The 'autorej' function of EEGLAB v12 was used to identify ICs with extreme values. Each component with activity above a rejection threshold of values  $\pm 50$  standard deviation was tagged as a possible artefact and the contribution it gave to that subject's EEG was removed. This resulted in a mean of 4.5 ICs being removed (range of 1–12). Visual classification using this IC-cleaned scalp EEG data is shown in Fig. 8. Using ICA to remove the contribution of obvious artefacts from the scalp EEG data improved the accuracy on this 'IC cleaned scalp EEG' over that from scalp EEG alone ( $p < 0.0001$ , on means of 0.68 AUC and 0.65 AUC). Mean classification from a single selected ICA channel remained higher than IC cleaned scalp EEG data (0.70 AUC to 0.68 AUC), but this was not significant ( $p = 0.1$ ). From this, we conclude that selected ICA data or selected IC cleaned scalp EEG data gives better accuracy in classifying vision from EEG data using this kind of SVM protocol.

### 3.7. Assessing accuracy over time with a sliding analysis window

For the previous results, classification was performed using data from the first 500 ms after each object was shown – from  $t = 0$  ms to  $t = +500$  ms, relative to the object appearing on the screen. In order to compare data within this time period, a 'sliding window' approach (Maris and Oostenveld, 2007) was then used, with windows of 500 and 50 ms. We use this sliding window approach to

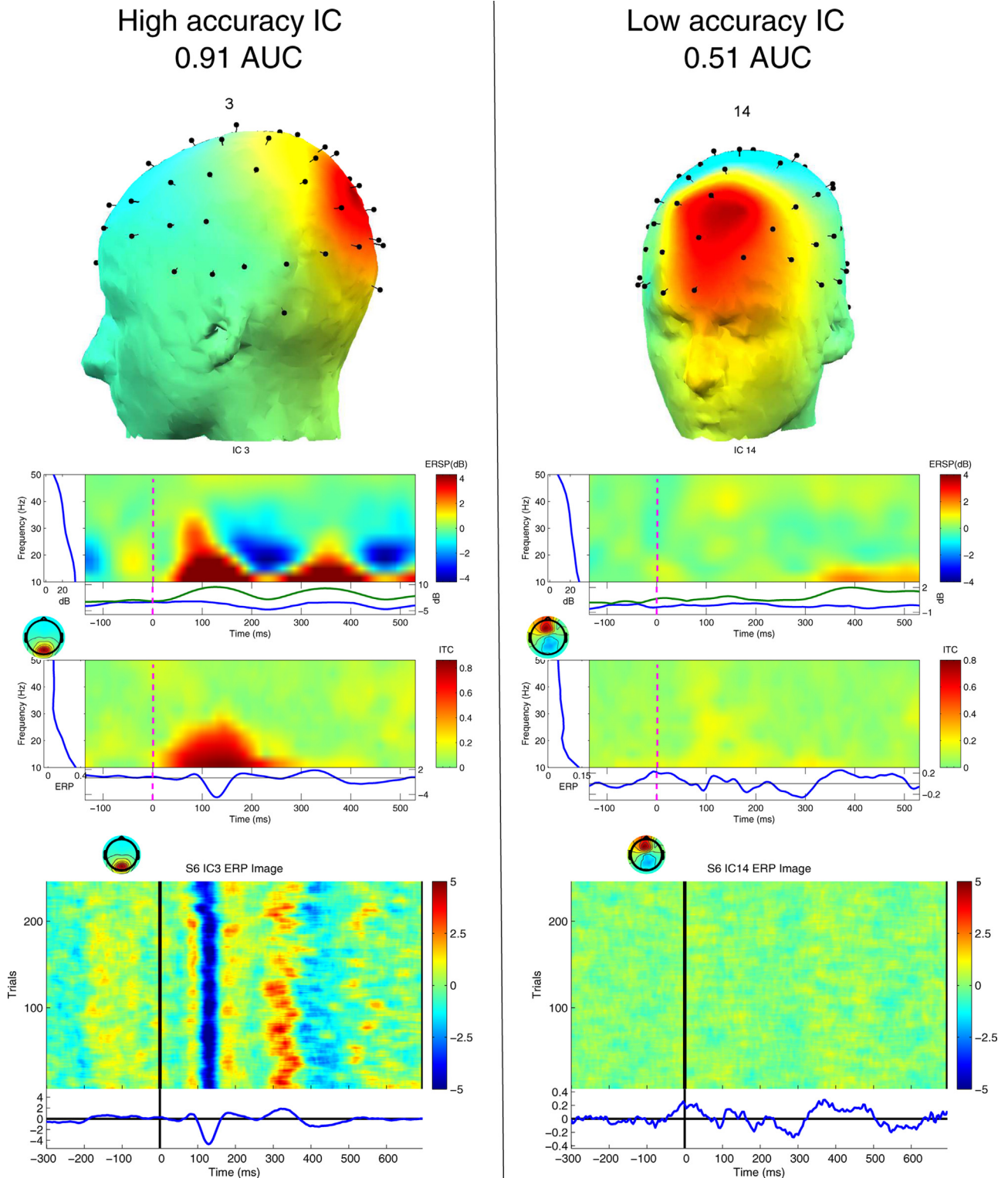
test the sensitivity of our approach to both the start of the data window, and the amplitude of the data window.

This sliding window protocol was performed in two conditions – first with the length of the analysis window remaining fixed at 500 ms, as before, and also with a narrower temporal window of 50 ms. At a window length of 500 ms, classification performance of scalp EEG data and IC data could be assessed when 0–500, 25–525, 50–550 ms, and so on up to 500–1000 ms, were used as positive training data. This data is shown in Fig. 9. Data from the one selected IC (black circles) gave highest accuracy at 0.72 AUC when using data from the interval within 25–525 ms. The one selected IC gave classification performance higher than other data sources at all time window intervals, greatly so until 100–600 ms, although performance drops steadily to around 0.54 AUC at 400–900 ms. Data from selected EEG channel follows this classification performance, but at around 0.1 AUC lower than IC data at 25–525 ms.

A smaller analysis window allows more precise timing resolution. At a window length of 50 ms, classification performance of scalp EEG data and IC data could be assessed when 0–50, 25–75, 50–100 ms, and so on up to 500–550 ms, were used as positive training data. This is shown in Fig. 10. Here, peak accuracy is no longer in the earliest intervals after the object appears, but instead in the slightly later intervals of 75–125 or 100–150 ms. Selected IC data has slightly higher average accuracy than selected EEG channel data. Selected EEG data has as peak accuracy at 0.65 AUC at the 100–150 ms interval. With training data consisting of 20 single-trials of just 50 ms, we find above-chance classification and the accuracy is considerably higher using data from 75 to 175 ms into the trial than at other intervals.

With only a tenth of the data (that from 50 ms rather than 500 ms), peak classification accuracy does drop from 0.72 AUC using 25–525 ms to 0.65 AUC using 100–150 ms. These data support conclusions that much of the task related EEG variance occurs in the 100–150 ms time interval, as models trained on that period have 0.65 AUC object classification accuracy. This is in agreement with related work showing a peak in 'decodability' at this time interval (Carlson et al., 2013), who also report this in

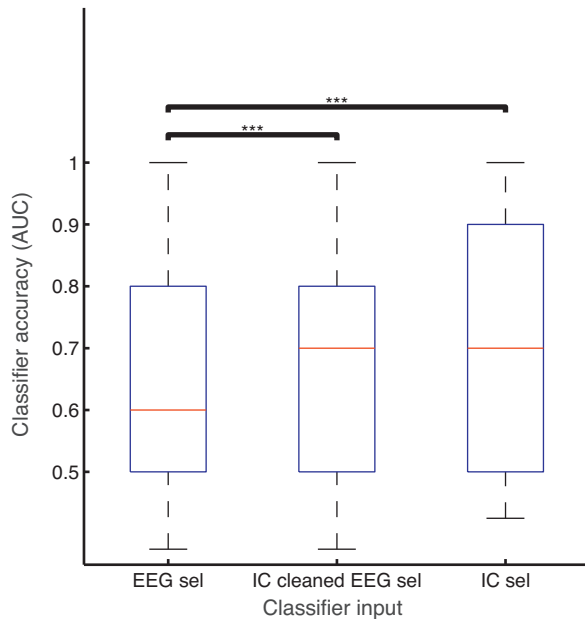




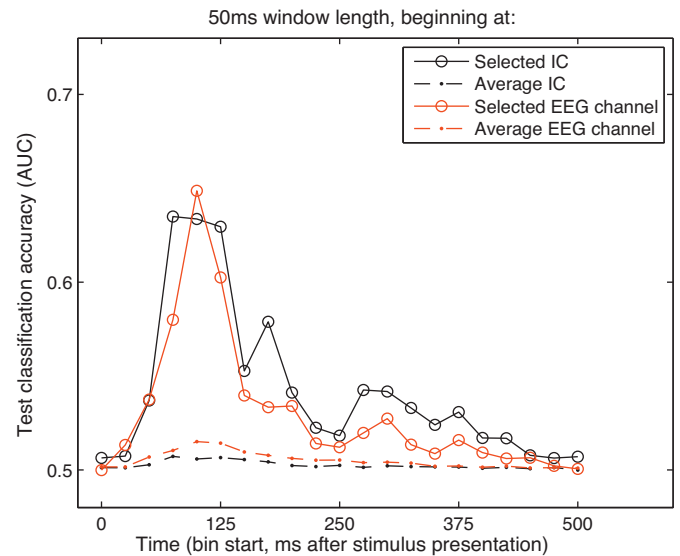
**Fig. 7.** Individual IC activation details. Activations of two example ICs in subject 6 – IC 3 (left) and IC 14 (right). **Top** standard 3D head models with EEG electrode positions as black pins and colours showing the spatial distribution of two ICs. In this example, IC 3 is an occipital component and IC 14 is a frontal component. **Mid** EEGLAB displays of frequency dynamics (event-related spectral perturbation) and plots of inter-trial coherence. **Bottom** corresponding IC ERP-image plots showing the 250 individual object presentation trials, with red indicating higher activation of this IC at that time. The lower plot is the standard IC ERP averaged over 250 trials. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

MEG object category discrimination. The results here agree with Carlson et al., 2013 in that object decoding performance peaks at around 100–150 ms, and that even later data from 375–425 ms gives higher object decoding performance than the early 25–75 ms

interval. The sliding window data above shows average classification performance across all seven subjects, using data from different EEG channels or ICs. Sliding window results separated for each subject and data source are shown for three random subjects



**Fig. 8.** Trial classification performance using EEG data, EEG data in which noisy ICs have been removed, and IC data. The box plot shows median accuracy of each data source in red, with the edges of the blue box indicating 25th and 75th percentiles. Asterisks (\*\*\*) indicate  $p < 0.0001$ , as found using a Wilcoxon signed-rank test. An average of 4.5 (of around 49 ICs total) were removed for the 'IC cleaned EEG data'. The increase in the median accuracy in the IC-cleaned EEG data suggests that these few artefactual ICs hindered ability to classify trials from scalp EEG data. The single selected IC is 'isolated' from the contribution from the artefactual IC noise. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 10.** The 'sliding window' is repeated, with a window length reduced to 50 ms of data is used for training models rather than 500 ms. We can see that data from 75 to 125 ms is sufficient to give 0.64 AUC accuracy. Accuracy falls off sharply around this period. Data from 0–50 ms and 25–75 ms (the first two time bins) are all at chance, suggesting no useful information for our classifiers at this time. Data from after 400 to 450 ms are also back at chance.

in Fig. 11. For each subject, two plots show colour-coded AUC for each electrode and each IC over the different latencies. High accuracy is concentrated in a few ICs, and those time bins starting before 150 ms appear to give higher accuracy.

### 3.8. Classification performance on each image

We report the success rate of classifying each object in Fig. 12. This shows the classification performance on each of the 50 objects used as visual stimuli using the selected EEG and IC data, averaged across the seven subjects. Highest accuracy was obtained when using the selected IC data on the 'log.jpg' with 0.89 AUC.

## 4. Discussion

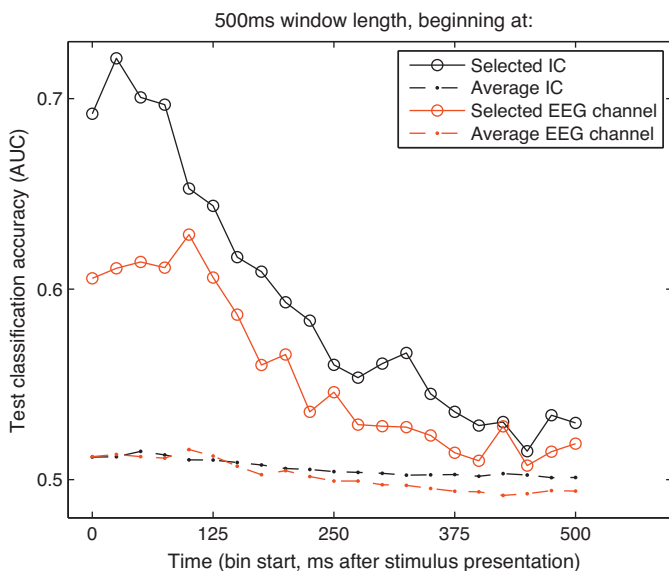
We detail a method for classifying visual state from EEG using machine learning. We demonstrate that this can distinguish data from visual object presentation trials from data without object presentation at 87% accuracy.

We found single-trial visual classification accuracy well above chance when using data from a single selected EEG channel or IC, with selected IC giving higher accuracy.

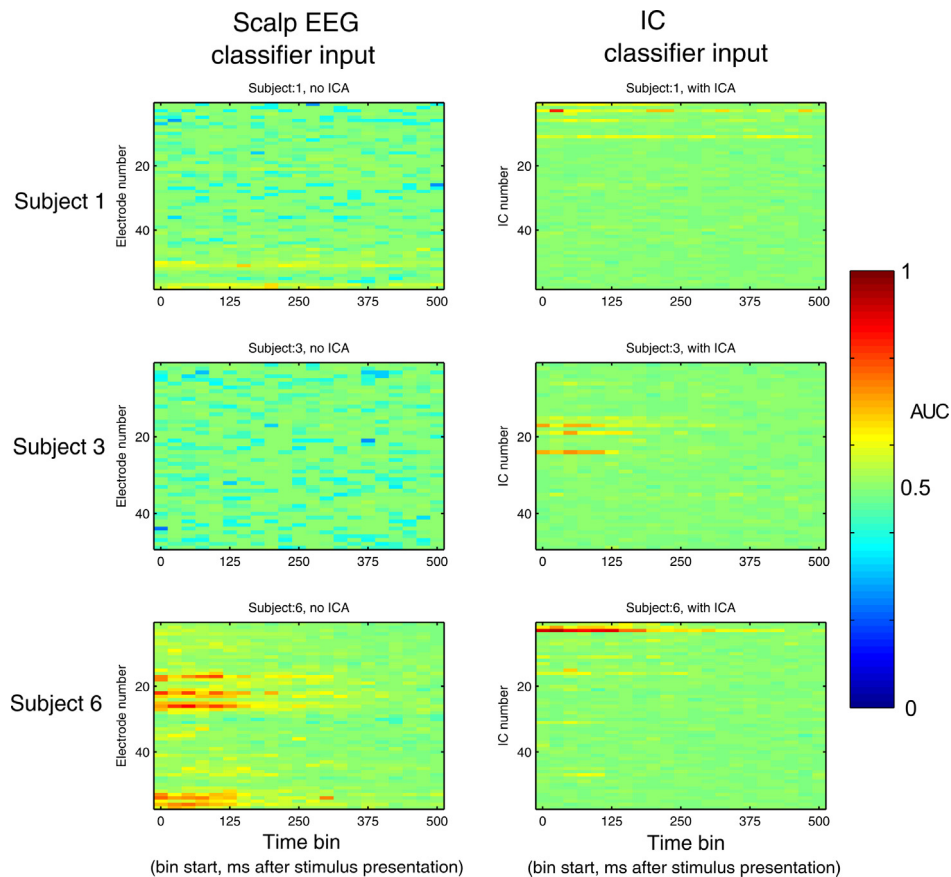
Further, we report single channel accuracy at 75–125 ms gives higher accuracy than other time bins, when classifying with these selected ICs (Fig. 10). This scoring of classifier accuracy automatically identifies electrodes, ICs, and time periods with activity that may be more relevant for the given trial.

### 4.1. Improving accuracy with selected input

The finding that input of a subset of a single selected data channel outperformed using all channels together deserves consideration. This may be due to classifier 'overfitting' – where the model parameters fit the properties of the training data too rigidly and so do not best generalise to classifying new data (Babyak, 2004). The overfitting when using all data could be due to suboptimal model parameters of the training error cost parameter 'C' and the radial-basis function kernel parameter 'γ'. However, these were the best values provided by a parameter search.



**Fig. 9.** A 'sliding window' analysis, to assess accuracy over trial time. Accuracy using data from 21 different time windows were used, all with length of 500 ms, first starting at 0–500 ms, and stepping through the trial times in 25 ms increments. Peak accuracy occurs here when using data from the selected IC (black circles), and models are trained to classify based on data from 25 to 525 ms after the object appears on screen, with 0.72 AUC. At the later time bin starting at 325 ms, the selected IC drops to 0.56 AUC, suggesting less trial-predictive information there, but is still above chance.



**Fig. 11.** A heatmap plot of the classifier performance (AUC) for a random three subjects using every electrode (left) and IC (right) as classifier input data, when using data from different timepoints in the trial. Outside the first 150 ms, accuracy is low. Peak accuracy is higher in IC data, with our classifier better able to separate classes in that data type. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Generalisation was improved by using appropriate dimensions of the classifier input data. Using one selected channel of input data rather than all channels or typical single channel gave increased mean accuracy. Notably, this considerably increased reliability (from 0.51 AUC to 0.65 AUC), with classification accuracy well above chance. This could be as a result of our implementation, but does indeed still show that a single selected channel of EEG input data can be a consistent and concise source of input data for high-accuracy classification.

#### 4.2. The use of ICA

When classifying presence of visual images, using classifier input of the IC activations from a single selected IC gave highest accuracy in all seven subjects at 500 ms window. Although the average IC input classification was approximately equal to the average single EEG channel input, those ICs that performed best on the initial 10 objects continued to outperform EEG channel input in further test data (0.70 AUC using selected IC data, compared to 0.65 AUC in the selected EEG channel). This corresponds to a significant increase in timepoints correctly labelled.

In trying to improve performance of neuroimaging classifiers, there are two domains commonly focussed upon: feature extraction and the classification itself (Farquhar and Hill, 2006). EEG data is intrinsically noisy, contains highly correlated features and has much variance both between different subjects and within the same subject over time. This suggests that considering decomposition of possible EEG sources could be useful when assessing features to extract from EEG (van Gerven et al., 2009). ICA does this

by decorrelating the inputs and attempting to minimise mutual information in forming the components.

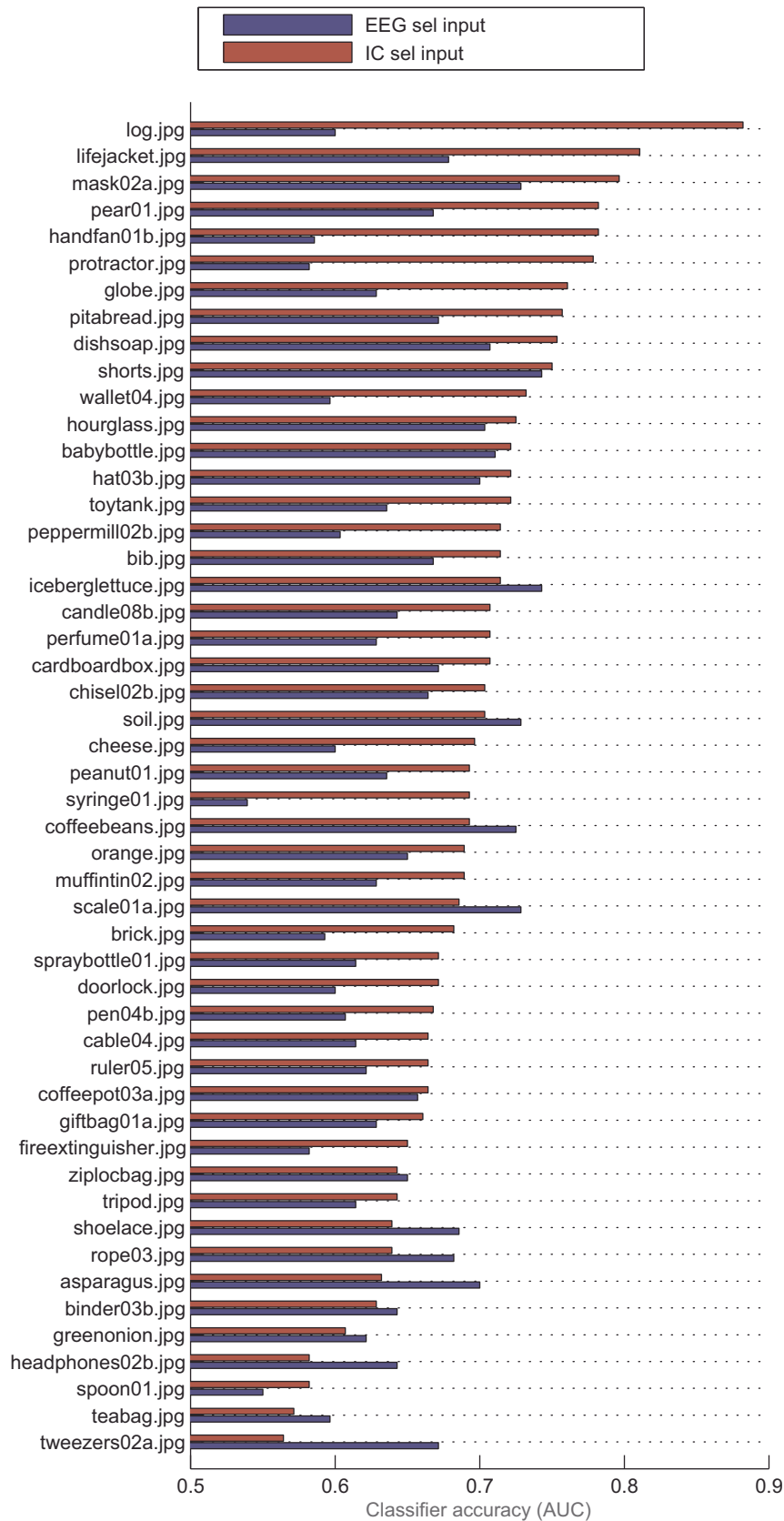
In previous studies, ICA has been shown to improve performance of a classifier in a simple auditory task (Hill et al., 2004). To our knowledge, the present report is the first to document ICA giving such an improvement in a visual object classification task.

For five of seven subjects, the single selected IC was found in occipital regions (Fig. 6). As expected, artefact related ICs had low predictive power to discriminate stimuli. Several of the components seemed to be related to eye movements, muscle artefacts or electrical noise, as identified by back-projected location and activation properties (Delorme et al., 2007). Nonetheless, we have shown here that several ICA components have a greater degree of information content that can be used to predict perceptual processes than unprocessed EEG data had.

ICA appeared to separate sources of noise into some ICs, and sources of task-related neural activity into other ICs. This resulted in the single-trial classification performance in many ICs being low, but a few ICs giving higher accuracy than that of any EEG channel. Thus, ICA may be thought of as concentrating source signals into de-noised ICs (as shown in the fewer, but more peaky, bands of yellow and red in Fig. 11), and so our SVM classifiers can give higher accuracy using this input. Examining IC input also has the advantage of assessing a source-space deconvolution (Bell and Sejnowski, 1995).

This view was also supported by an improvement in trial classification accuracy from EEG data where probable-artefacts have been removed (see Fig. 8).

Within subjects, many EEG channels had relatively similar predictive power. While the magnitude of the averaged ERP might be quite different on electrodes across the scalp, at a single-trial



**Fig. 12.** Classification performance shown by each object of our 50 visual object stimuli, sorted by IC AUC. This is averaged across the seven subjects, using the selected IC data (shown in dark red) and selected EEG channel data (dark blue). Single-trials with pictures of the log, life-jacket and mask were labelled correctly at high accuracy, and the IC outperformed scalp electrode data for most objects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



level much of the signal present at any electrode is also present on neighbouring electrodes (Luck, 2005) albeit at a different scale. As the SVM classifier here is separating classes in a high-dimensional hyperplane, it can remain sensitive to small (but consistent) separating features. Thus, features that would not be obviously visually apparent on a plotted ERP can still be captured.

We suggest that the task predictive power of an IC has the potential to inform theories of visual cognition. For example, an IC that is localised in neck muscles and also gives low accuracy when used to classify the task, would be an unlikely target for further investigation. In contrast, an IC that is spatially localised in visual cortex and that gives high accuracy of classifying the visual state under specific conditions may be a fruitful target to profile.

#### 4.3. Visual object processing

Although EEG studies have shown stimuli-specific ERP separation within 150 ms (Thorpe et al., 1996; Johnson and Olshausen, 2005; Mouchetant-Rostaing and Giard, 2000), there appears to be some uncertainty whether these differences represent low-level visual properties or higher-level cognitive categorisation of visual stimuli. Related work in macaque monkeys using multi-unit neuron cell recordings seems to indicate clear object- and category-specific information in primate inferior temporal cortex at 100–125 ms after presentation (Hung et al., 2005). Cell firing was used as input for a classifier to correctly identify the presented object stimulus at 70% accuracy using only the spiking behaviour of a few neurons in inferior temporal cortex.

This neuron spiking data strongly indicates that the brain activity is present at these time-scales, but that, of course, does not necessitate that we could also observe this within EEG data, which would allow more probing in humans. Clearly, the ability of averaged ERP EEG analysis alone is limited in addressing this (Makeig et al., 2004; Rousselet and Pernet, 2011). We suggest that our approach of ICA and SVM machine learning is a more suitable one.

#### 4.4. Timing of accurate trial classification throughout a trial

With use of the sliding window moving through data from different periods of the time into a trial (Figs. 9 and 10), we found that data from the time period at 75–150 ms was most useful for classifier performance. This is in agreement with studies of ERPs in visual object presentation (Thorpe et al., 1996), where the difference in the ERP from different objects is small before 75 ms and after 350 ms.

This profile of accuracy using these selected occipital ICs over time is in agreement with related studies of early visual object processing listed above, with the current procedure automatically finding these task-relevant ICs.

We suggest that this reveals how the timing of relevance of the information contained in different EEG channels, processing methods and IC sources can be assessed by how well they can be used to predict experimental conditions with these kinds of SVM classifiers.

#### 4.5. Cognitive implications and future work

The identification of EEG sources that have robust task-related predictive power would allow cognitive science experimental designs that target that specific IC. Getting a profile of IC localisation, task dependence, temporal and spectral properties may be much more accessible in probing possible underlying neural processes in greater detail than using EEG data alone. Observing increased accuracy when using IC activation input suggests that they are a promising target, although the extent to which specific ICs might relate to underlying neural processes is not yet well known (Onton et al., 2006).

While standard ERP components such as the N170 and P300 are frequently reported landmarks in EEG data for visual perception (Joyce and Rossion, 2005) and oddball responses (Polich and Kok, 1995), such ERPs are limited in being averages of electrode traces, and so may lose useful information (Rousselet and Pernet, 2011).

Other components have been suggested as descriptions for EEG data. Philiastides and Sajda, 2006 report their generation of components to best decode trials in which either cars or faces are shown, and find one such component that resembles properties of the N170. This demonstrates an alternate description that may underlie these ERP components, and so give another way of studying them.

In the current study, we also report components that give high decoding performance. Here ICA was used to generate the ICs, and so represents possible sources of generated EEG activity, obtained blindly from the data. The automated scoring procedure we used identified a group of occipital IC components that represent the best data sources for classifying visual object presence in single trials.

Other studies of visual perception have considered the similarity of fMRI activity in inferior temporal cortex in response to objects of different categories (Kriegeskorte et al., 2008). The current study used 50 images from the BOSS image database (Brodeur et al., 2010), all of which are likely to be considered 'natural objects' or 'artificial objects' in category. We studied a range of single images and did not examine object categories, but future studies could target discrimination of categories with relevant training stimuli.

Fast, high accuracy classification of visual state also allows real-time detection within a recording session, and so experiments involving EEG feedback or manipulation could be performed. Do any ICs have activations associated with borderline visual percepts, or is the task-related representation we see here downstream of an all-or-nothing percept trigger?

These components – along with our SVM prediction models – can classify new trials of visual stimulation well, but could likely be used to probe other steps of visual processing and perception. We suggest that profiling the activity of these components identified by ICA in different tasks, quantified with machine learning, might be fruitful conceptual targets for future visual processing experiments. Might any of their classifiers respond to imagined objects rather than presented objects? Given several objects presented simultaneously to a subject, might some classifiers respond to the attention of the subject and some respond only to the immediate visual input, regardless of context? With a visual task of noisy stimuli where object perception is only sometimes reported, will some classifiers predict onset of perception even before perception occurs?

#### 4.6. Limitations

The metric used for 'prediction power' gives an indication of task-related information, but it will also have some dependencies on the properties of the classifier used (Meyer et al., 2003). Nonetheless, the prediction power of those ICs that likely capture artefactual signal is low, suggesting that task prediction power can be a valid proxy for task-related information content.

Models using data from all channels simultaneously (as in Section 3.1) gave low average accuracy. This suggests this classifier model was not fitting the signal-to-noise of this data well, and overfitting (Chang and Lin, 2011). Using one selected 'best' IC resolved this, but at the expense of not utilising all available input data.

It may be the case that ICA was particularly useful here as little other preprocessing was used. Prior research suggests that use of tailored data extraction can increase accuracy on classifying EEG data (Blankertz et al., 2006), where frequencies of interest, channels of interest and input ranges are manually specified. We suggest that in cases where the anticipated activity is relatively unknown, use of the more automated ICA might avoid misuse of tailored data extraction when it is not appropriate.

We considered classifying the presence or absence of single images. It could be possible to extend this to instead classify each image against the other images: as 'image1 vs image2' rather than 'image1-present' versus 'image1-not-present', or classifying multiple images simultaneously through use of a multi-class classifier, but we have not addressed this here.

#### 4.7. Conclusion

We presented a method for automatically scoring task-related information present within EEG transforms using SVM classifiers trained on that data. We found 'object presence' was classified at 0.70 AUC (87%) when using data from a single selected IC. Further, the use of a sliding window analysis revealed that the time window spanning 75–150 ms gave the highest accuracy, when using these selected ICs. We suggest that this method of machine learning and independent 'source separation' might allow detailed probing of information content within EEG data.

#### Acknowledgements

This work was supported in part by grants EP/F500385/1 and BB/F529254/1 for University of Edinburgh Neuroinformatics, from EPSRC, BBSRC and MRC.

#### References

- Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med* 2004;66(3):411–21.
- Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 1995;7(6):1129–59.
- Blankertz B, Müller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlögl A, et al. The BCI competition. III: validating alternative approaches to actual BCI problems. *IEEE Trans Neural Syst Rehabil Eng* 2006;14(2):153–9.
- Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR. Optimizing spatial filters for Robust EEG single-trial analysis. *IEEE Signal Process Mag* 2008;25(1):41–56.
- Brainard DH. The psychophysics toolbox. *Spat Vis* 1997;10(4):433–6.
- Brodeur MB, Dionne-Dostie E, Montreuil T, Lepage M. The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One* 2010;5(5).
- Burges C. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 1998;167:121–67.
- Carlson T, Tovar DA, Alink A, Kriegeskorte N. Representational dynamics of object vision: the first 1000 ms. *J Vis* 2013;13(10):1–19.
- Chang C, Lin C. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011;2(3):1–27.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273–97.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004;134(1):9–21.
- Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 2007;34(4):1443–9.
- Donoghue JP. Bridging the brain to the world: a perspective on neural interface systems. *Neuron* 2008;60(3):511–21.
- Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am Stat* 1983;37(1):36–48.
- Farquhar J, Hill J. In: Learning optimal EEG features across time, frequency and space, in NIPS 2006 Workshop Current Trends Brain-Computer Interfacing; 2006. <http://www.kyb.mpg.de/publications/attachments/nips2006.4262%5B0%5D.pdf>
- Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000;16(10):906–14.
- Gaspar CM, Rousselet GA, Pernet CR. Reliability of ERP and single-trial analyses. *Neuroimage* 2011;58(2):620–9.
- van Gerven M, Farquhar J, Schaefer R, Vlek R, Geuze J, Nijholt A, et al. The brain-computer interface cycle 2009;6(4):041001.
- Hand DJ. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach Learn* 2009;77(1):103–23.
- Hill N, Lal T, Bierig K. Attention modulation of auditory event-related potentials in a brain-computer interface. In: *IEEE Int Work Biomed Circuits Syst*; 2004. p. 3–6.
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. *Science* 2005;310(5749):863–6.
- Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Networks* 2000;13(4–5):411–30.
- Johnson J, Olshausen B. The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J Vis* 2005;5(4):299–312.
- Joyce C, Rossion B. The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. *Clin Neurophysiol* 2005;116(11):2613–31.
- Keerthi S, Lin C. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput* 2003;15(7):1667–89.
- Kleiner M, Brainard D, Pelli D, Ingling A. What's new in Psychtoolbox-3. In: *Percept ECVF Abstr Suppl*; 2007. p. 14.
- Kriegeskorte N, Mur M, Ruff Da, Kiani R, Bodurka J, Esteky H, et al. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 2008;60(6):1126–41.
- Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng* 2007;4(2):R1–13.
- Luck S. An introduction to the event-related potential technique. Cambridge, MA: MIT Press; 2005.
- Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. *Trends Cogn Sci* 2004;8(5):204–10.
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 2007;164:177–90.
- Mason S, Graham N. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Q J R Meteorol Soc* 2002;128(584):2145–66.
- Meyer D, Leisch F, Hornik K. The support vector machine under test. *Neurocomputing* 2003;55(1–2):169–86.
- Mouchetant-Rostaing Y, Giard M. Early signs of visual categorization for biological and non-biological stimuli in humans. *Cogn Neurosci* 2000;11(11):2521–5.
- Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *J Neurosci Methods* 2008;167(1):82–90.
- Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. *Neurosci Biobehav Rev* 2006;30(6):808–22.
- Osuna E, Freund R, Girosi F. Training support vector machines: an application to face detection. In: *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*; 1997. p. 130–6.
- Philastides MG, Sajda P. Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb Cortex* 2006;16(4):509–18.
- Polich J, Kok A. Cognitive and biological determinants of P300: an integrative review. *Biol Psychol* 1995;41(2):103–46.
- Rousselet GA, Husk JS, Bennett PJ, Sekuler AB. Single-trial EEG dynamics of object and face visual processing. *Neuroimage* 2007;36(3):843–62.
- Rousselet GA, Pernet CR. Quantifying the time course of visual object processing using ERPs: it's time to up the game. *Front Psychol* 2011;2:1–6.
- Sajda P, Gerson A, Müller KR, Blankertz B, Parra L. A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *IEEE Trans Neural Syst Rehabil Eng* 2003;11(2):184–5.
- Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature* 1996;381(6582):520–2.
- Wang Y, Jung TP. Improving brain-computer interfaces using independent component analysis. In: Allison BZ, Dunne S, Leeb R, Del R, Millán J, Nijholt A, editors. *Toward. Pract. Brain-Computer Interfaces*. Berlin, Heidelberg: Springer; 2012. Biological and Medical Physics, Biomedical Engineering.
- Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, et al. Brain-computer interface technology: a review of the first international meeting. *IEEE Trans Rehabil Eng* 2000;8(2):164–73.

# Bibliography

- Andersson, P., Pluim, J. P. W., Siero, J. C. W., Klein, S., Viergever, M. a., and Ramsey, N. F. (2011). Real-time decoding of brain responses to visuospatial attention using 7T fMRI. *PloS one*, 6(11):e27638.
- Artamonov, O. (2004). Contemporary LCD Monitor Parameters and Characteristics. *Xbit Labs webpage*, (<http://goo.gl/tqznH2>).
- Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3):411–421.
- Barber, D. (2011). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.
- Beckmann, C. F., Mackay, C. E., Filippini, N., and Smith, S. M. (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *Human Brain Mapping Conference*, page 181.
- Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–59.
- Biosemi (2013). [www.biosemi.com/software\\_biosemi\\_analysis.htm](http://www.biosemi.com/software_biosemi_analysis.htm).
- Blankertz, B., Müller, K.-R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., Pfurtscheller, G., Millán, J. D. R., Schröder, M., and Birbaumer, N. (2006). The BCI competition. III: Validating alternative approaches to actual BCI problems. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 14(2):153–9.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K.-r. (2008). Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine*, 25(1):41–56.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, (10):433–436.
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., and Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5(5).

- Bullinaria, J. a. and Levy, J. P. (2013). Limiting Factors for Mapping Corpus-Based Semantic Representations to Brain Activity. *PLoS ONE*, 8(3).
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 167:121–167.
- Buzsáki, G., Anastassiou, C. a., and Koch, C. (2012). The origin of extracellular fields and currents EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420.
- Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R. Q., Koch, C., and Fried, I. (2010). On-line, voluntary control of human temporal lobe neurons. *Nature*, 467(7319):1104–8.
- Chan, A. M., Halgren, E., Marinkovic, K., and Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, 54(4):3028–3039.
- Chang, C. and Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27.
- Chih-Wei Hsu, Chih-Chung Chang and Lin, C.-J. (2008). A Practical Guide to Support Vector Classification. *BJU international*, 101(1):1396–400.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.
- Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., and Makeig, S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Computational intelligence and neuroscience*, 2011:130714.
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PloS one*, 7(2):e30135.
- Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4):1443–9.
- Donoghue, J. P. (2008). Bridging the brain to the world: a perspective on neural interface systems. *Neuron*, 60(3):511–21.
- Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48.

- Farnell (2012). <http://uk.farnell.com/centronic/bpx65rt/photodiode/dp/548789>.
- Farquhar, J. and Hill, J. (2006). Learning optimal EEG features across time, frequency and space. *Advances In Neural Information Processing Systems*.
- Furey, T., Cristianini, N., and Duffy, N. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914.
- Gaspar, C. M., Rousselet, G. A., and Pernet, C. R. (2011). Reliability of ERP and single-trial analyses. *NeuroImage*, 58(2):620–629.
- Gruener, W. (2008). <http://www.tgdaily.com/business-and-law-features/36119-lcd-tvs-outship-crt-tvs-for-the-first-time>.
- Hammon, P. and Makeig, S. (2008). Predicting reaching targets from human EEG. *Signal Processing ...*, (January 2008):69–77.
- Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1):103–123.
- Hill, N., Lal, T., and Bierig, K. (2004). Attention modulation of auditory event-related potentials in a brain-computer interface. In *IEEE International Workshop on Biomedical Circuits and Systems*, pages 3–6.
- Horovitz, S. G., Skudlarski, P., and Gore, J. C. (2002). Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: a parametric approach to combining fMRI and ERP. *Magnetic resonance imaging*, 20(4):319–25.
- Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–6.
- Hyvärinen, A. (2011). *Testing the ICA mixing matrix based on inter-subject or inter-session consistency.*, volume 58.
- Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–30.
- Itti, L. and Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(March):194–203.
- Johnson, J. and Olshausen, B. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, 5(4):299–312.
- Joyce, C. and Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. *Clinical Neurophysiology*, 116(11):2613–31.

- Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185):352–5.
- Keerthi, S. and Lin, C. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15(7):1667–1689.
- Kirchner, H. and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, 46(11):1762–76.
- Kleiner, M., Brainard, D., Pelli, D., and Ingling, A. (2007). What’s new in Psychtoolbox-3. In *Perception ECVF Abstract Supplement*, pages 433–436.
- Koch, C. (2004). *The Quest for Consciousness*. New York.
- Kriegeskorte, N., Mur, M., Ruff, D. a., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P. a. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–41.
- Kronegg J., V. S. P. T. (2005). Analysis of bit-rate definitions for Brain-Computer Interfaces. *Proceedings of the 2005 International Conference on Human-Computer Interaction, HCI’05*, pages 40–46.
- Langlois, D., Chartier, S., and Gosselin, D. (2010). An Introduction to Independent Component Analysis : InfoMax and FastICA Algorithms. *Tutorials in Quantitative Methods for Psychology*, 6(1):31–38.
- Li, Y., Long, J., Yu, T., Yu, Z., Wang, C., Zhang, H., and Guan, C. (2010). An EEG-based BCI system for 2-D cursor control by combining Mu/Beta rhythm and P300 potential. *IEEE Transactions on Biomedical Engineering*, 57(10):2495–2505.
- Lotte, F., Congedo, M., Lécuyer, a., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, 4(2):R1–R13.
- Luck, S. (2005). *An introduction to the event-related potential technique*. MIT Press, Cambridge, Massachusetts.
- Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8(5):204–10.
- Makeig, S., J. Bell., A., Jung, T.-P., and Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems*, 8:145–151.
- Marr, D. (1982). *”Vision. A Computational Investigation into the Human Representation and Processing of Visual Information.”*.

- Mason, S. and Graham, N. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, 125(584):2145–2166.
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(9):2725–32.
- Meyer, D., Leisch, F., and Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, 55(1-2):169–186.
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Kai-Min, C., Malave, V. L., Mason, R. a., and Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, 320(January):1191–1195.
- Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-a., Morito, Y., Tanabe, H. C., Sadato, N., and Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–29.
- Mouchetant-Rostaing, Y. and Giard, M. (2000). Early signs of visual categorization for biological and non-biological stimuli in humans. *Cognitive Neuroscience*, 11(11):2521–5.
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain : a Journal of Neurology*, pages 701–22.
- Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1):82–90.
- Murphy, B., Baroni, M., and Poesio, M. (2009). EEG responds to conceptual stimuli and corpus semantics. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2(August):619–627.
- Ng, A. (2009). Advice for applying machine learning. *CS229 Class Notes*.
- Niedermeyer, E. and Da Silva, F. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.
- Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J. (2011). Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Current Biology*, pages 1641–1646.

- Noachtar, S. and Rémi, J. (2009). The role of EEG in epilepsy: a critical review. *Epilepsy & behavior : E&B*, 15(1):22–33.
- Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567.
- Oja, E. and Yuan, Z. (2006). The FastICA Algorithm Revisited: Convergence Analysis. *Neural Networks, IEEE Transactions on*, 17(6):1–20.
- Onton, J., Delorme, A., and Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, 27(2):341–56.
- Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, 30(6):808–22.
- Osuna, E., Freund, R., and Girosit, F. (1997). Training support vector machines: an application to face detection. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130–136.
- Philiastides, M. G. and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, 16(4):509–18.
- Polich, J. and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, 41(2):103–46.
- Rousselet, G. A. (2012). Does Filtering Preclude Us from Studying ERP Time-Courses? *Frontiers in psychology*, 3(May):131.
- Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2007). Single-trial EEG dynamics of object and face visual processing. *NeuroImage*, 36(3):843–62.
- Rousselet, G. A. and Pernet, C. R. (2011). Quantifying the time course of visual object processing using ERPs: it’s time to up the game. *Frontiers in Psychology*, 2:1–6.
- Sajda, P., Gerson, A., Müller, K.-R., Blankertz, B., and Parra, L. (2003). A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitative Engineering*, 11(2):184–5.
- Scalaidhe, S. P., Wilson, F. a., and Goldman-Rakic, P. S. (1999). Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cerebral cortex (New York, N.Y. : 1991)*, 9(5):459–75.
- Smith, M. L., Gosselin, F., and Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology : CB*, 22(3):191–6.



- Stewart, A. X., Nuthmann, A., and Sanguinetti, G. (2014). Single-trial classification of EEG in a visual object task using ICA and machine learning. *Journal of Neuroscience Methods*, 228:1–14.
- Swartz, B. E. (1998). The advantages of digital over analog recording techniques. *Electroencephalography and clinical neurophysiology*, 106(2):113–7.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.
- Thut, G., Miniussi, C., and Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Current biology : CB*, 22(16):R658–63.
- Tsotsos, J. K. and Rothenstein, A. (2011). Computational models of visual attention. *Scholarpedia*.
- van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., and Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4).
- Vanrullen, R. and Thorpe, S. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*.
- Vigario, R., Sarela, J., Jousmiki, V., Hamalainen, M., and Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *Biomedical Engineering, IEEE Transactions on*, 47(5):589–593.
- Wang, D., Miao, D., and Blohm, G. (2012). Multi-class motor imagery EEG decoding for brain-computer interfaces. *Frontiers in Neuroscience*, 6(October):1–13.
- Wang, P. and Nikolic, D. (2011). An LCD monitor with sufficiently precise timing for research in vision. *Frontiers in Human Neuroscience*, 5(August):85–85.
- Wang, Y. and Jung, T.-P. (2012). Improving brain-computer interfaces using independent component analysis. In Allison, B. Z., Dunne, S., Leeb, R., Del R. Millán, J., and Nijholt, A., editors, *Towards Practical Brain-Computer Interfaces*, Biological and Medical Physics, Biomedical Engineering, pages 67–83. Springer Berlin Heidelberg.
- Weber, M. J. and Thompson-Schill, S. L. (2010). Functional neuroimaging can support causal claims about brain function. *Journal of Cognitive Neuroscience*, 22(11):2415–6.
- Widmann, A. and Schröger, E. (2012). Filter effects and filter artifacts in the analysis of electrophysiological data. *Frontiers in psychology*, 3(July):233.

- Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., Donchin, E., Quatrano, L. A., Robinson, C. J., and Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions On Rehabilitation Engineering*, 8(2):164–173.
- Zhang, X., Zhaoping, L., Zhou, T., and Fang, F. (2012). Neural activities in v1 create a bottom-up saliency map. *Neuron*, 73(1):183–92.
- Zhaoping, L. (2014). *Understanding vision: theory, models, and data*.